Multi-Objective Optimization in Semi-Empirical Quantum Chemistry Methods: Applications to Liquid Water Simulations

Universität Zürich[UZH]

Dissertation
zur
Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)
vorgelegt der
Mathematisch-naturwissenschaftlichen Fakultät
der
Universität Zürich

von

Yannick B. Misteli

aus
Liestal BL

Promotionskomitee

Prof. Dr. Jürg Hutter (Vorsitz und Leitung)
Prof. Dr. Peter Hamm

Zürich, 2016

Es ist Unsinn, sagt die Vernunft
Es ist was es ist, sagt die Liebe
—Erich Fried

To my parents...

Acknowledgements

I would like to thank Professor Jürg Hutter for giving me the opportunity to do my PhD in his group in Zürich and let me develop on a professional and personal level.

I would also like to thank my colleges Mauro del Ben, Mandes Schönherr and Jan Wilhelm for all their support, discussions and the strength to endure my music in the office. Additionally, I would like to thank Ole Schütt and Christian Müller for the inspiring discussions and their profound knowledge in optimization and computer science.

Gratitude is also expressed towards Chris Mundy and Marcel Baer for a very pleasant stay at PNNL and a great collaboration with many enthusiastic discussions, support and encouragements.

Furthermore, I would like to thank all the present and former PostDocs and PhD students of my group for a great work atmosphere and wonderful time.

Finally, a special thanks to my parents and all the friends for their continuous support during the whole time. In particular, a great debt of gratitude is owed to my sister.

Zürich, 20 October 2015 Y. B. M.

Abstract

The objective of this thesis is to pave the way for a systematic parameter optimization in theoretical chemistry. On all levels of theory in computational chemistry, from ab inito to semi-empirical or force-field approaches, one is confronted with parameter optimization to some extend. The arising issue with parametrization is that usually the underlying physical model is supposed to predict not only one property such as total energy but also atomic structure, dipole or density. Therefore, these properties should be employed as objective functions in the optimization process, where the goal is to minimize these functions usually composed as error estimators of reference values. Inevitably, this leads to a multi-objective optimization problem and the common and most basic approach of scalarizing the respective objectives into a single-objective function has proven to be particularly problematic. The issues of the scalarization or preference-based approach are how the normalization and weighting is carried out besides the more fundamental problem that some functions can not be added up properly and non-convex solutions can not be recovered. These aspects make it necessary to introduce a more systematic and inherent approach. In this work we present a multi-objective algorithm based on stochastic optimization which is applied to semi-empirical quantum chemistry methods. The semi-empirical methods are based on the Hartree-Fock formalism, but introduce more approximations and hence are an order of magnitude faster than traditional ab inito methods. Although these methods are not as fast as the classical force-field approaches, they describe the system quantum mechanically and therefore are able to reproduce important effects such as forming and breaking bonds and polarizability. The combination of the accuracy of ab initio methods with the speed of the force field approach makes the semi-empirical methods a valuable tool to access larger systems and longer time-scales in molecular dynamics simulations without significant loss of accuracy.
In this work, we employ the multi-objective parameter optimization algorithm to the Neglect of the Diatomic Differential Overlap (NDDO) model, one of the most prominent representative thereof PM6, and Density Functional Tight Binding (DFTB). Applying these methods to liquid, bulk and interface simulations of water, we show that large improvements can be achieved when the optimized parameter set is employed. Thus, we establish a systematic and integral parameter optimization algorithm for multi-objective optimization in quantum chemistry. With the present work we laid the foundation for the application for systematic multi-objective optimization in computational chemistry, with further significance of force field development or basis set optimization.

Zusammenfassung

Ziel dieser Arbeit ist, den Weg für eine systematischere Herangehensweise der Parameter-Optimierung zu ebnen. Auf allen konzeptionellen Ebenen der computergestützten Chemie von ab-initio über semi-empirische Methoden zu Kraftfelder-Ansätzen, sieht man sich der Parameter-Optimierung ausgesetzt. Das Problem dabei ist, dass das unterliegende physikalische Modell nicht nur eine Quantität wie totale Energie reproduzieren soll, sondern zusätzlich auch Atomstrukturen, Dipole und Dichte. Demnach sollten diese Gegebenheiten in den Optimierungsprozess als Funktionen einfliessen. Diese Zielfunktionen bestehen meistens aus Fehlerfunktionen zu einem gegebenen Referenzwert. Unvermeidbar ist dabei, dass mehrere Zielfunktionen auftreten und üblicherweise wurde dieses Problem durch Skalarisierung, beziehungsweise durch Gewichten und Aufsummieren in eine einzige Zielfunktion überführt. Hierbei ist aber nicht klar, wie die Normalisierung und Gewichtung ausgeführt werden soll. Zusätzlich gibt es auch Fälle, in deren die Funktionen gar nicht addiert werden können. Zudem bleiben nicht konvexe Lösungen verborgen. Diese Aspekte machen es unumgänglich, einen systematischeren Ansatz zu wählen. In dieser Arbeit präsentieren wir einen multi-kriteriellen Optimierungs-Algorithmus, welcher im Bereich der semi-empirischen Quantenchemie angewandt wird. Die semi-empirischen Methoden basieren auf dem Hartree-Fock-Formalismus, führen aber zusätzliche Annäherungen ein und sind somit eine Grössenordnung schneller als traditionelle ab-initio Rechnungen. Obwohl diese Methoden nicht so schnell sind wie der Kraftfeld-Ansatz, beschreiben sie das System quantenmechanisch und ermöglichen deshalb den Beschrieb von Bildung und Bruch chemischer Verbindungen und Polarisation. Die Kombination der Genauigkeit der ab-initio-Methoden und die Geschwindigkeit des Kraftfeld-Ansatzes machen die semi-empirsche Vorgehensweisen zu einem wertvollen Instrument um grössere Systeme und längere Molekulardynamik-Simulationen zu ermöglichen. In dieser Arbeit wenden wir einen multi-kriteriellen Parameter-Optimierungsalgorithmus auf das Neglect of the Diatomic Differential Overlap (NDDO) Modell, beziehungsweise auf die Methode PM6, und Density Functional Tight Binding (DFTB) an. Diese Vorgehensweisen werden an Simulationen von flüssigem Wasser getestet. Wir konnten zeigen, dass eine erhebliche Verbesserung erzielt werden kann, wenn optimierte Parameter verwendet werden. Somit haben wir einen Weg aufgezeigt, wie Parameter-Optimierung mit mehreren Zielfunktionen systematischer als in der Vergangenheit ausgeführt werden kann. Zusätzlich ergeben sich Anwendungsbereiche in der Kraftfeld-Entwicklung oder der Basis-Satz Optimierung, die von diesem Ansatz profitieren könnten.

Contents

Acknowledgements **i**

List of figures **ix**

List of tables **xi**

Introduction **1**

1 Optimization **5**
1.1 Optimization Overview 5
1.2 Single and Multi-Objective Optimization 10
1.2.1 Fundamental Differences 10
1.2.2 Two Approaches to Multi-Objective Optimization 11
1.3 Multi-Objective Optimization 15
1.3.1 Basic Concepts 15
1.3.2 Difference with Single-Objective Optimization 16
1.3.3 Dominance and Pareto Optimality 17
1.3.4 Classical Methods 25
1.3.5 Evolutionary Algorithm and Covariance Matrix Adaptation 28
1.3.6 Multi-objective Covariance Matrix Adaptation Evolution Strategy 35

2 Semiempirical Methods **41**
2.1 Introduction 41
2.2 Historical Overview 43
2.3 Methods 45
2.3.1 Basic Concepts and Theory 45
2.3.2 Hückel Molcular Orbital Theory 51
2.3.3 Neglect of Diatomic Differential Overlap 52
2.3.4 DFTB 56

3 Applications **61**
3.1 Introduction 62
3.2 Mechanisms 62
3.2.1 Implementation 63

3.2.2 Visualization of n-dimensional Pareto front 64
3.2.3 Selection from n-dimensional Pareto sets 66
3.3 Multi-objective setup 67
3.4 Objective Functions 67
3.4.1 Root-Mean-Square-Deviation of atomic positions (RMSD) 68
3.4.2 Enthalpy Error 69
3.4.3 Charge Error 70
3.5 Reference Data 70
3.5.1 Water27 71
3.5.2 BEGDB water clusters 71
3.5.3 QCE clusters 74
3.6 Reparametrizing PM6 74
3.6.1 Parameters 76
3.6.2 Results 78
3.7 Parametrizing DFTB3 γ-function 84
3.7.1 γ-function and Parameters 84
3.7.2 Results 88
3.8 Bulk water properties 90
3.8.1 Computational Details 90
3.8.2 Radial Distribution Functions 91
3.8.3 Density 94
3.8.4 Surface tension 96
3.8.5 Diffusion 103

4 Summary and Outlook **107**
4.1 Summary 107
4.2 Outlook 109

Bibliography **129**

List of Figures

1.1 Global and local optima . . . 6
1.2 Gradient descent . . . 7
1.3 Serial-vs-parallel . . . 8
1.4 Piz Daint Supercomputer . . . 8
1.5 Taxonomy of algorithms . . . 9
1.6 Trade-off solutions . . . 11
1.7 The two approaches to multi-objective optimization . . . 12
1.8 Weighted sum introduction . . . 14
1.9 Mapping from decision space . . . 16
1.10 Concept of domination . . . 18
1.11 Pareto-front . . . 20
1.12 Pareto-level . . . 23
1.13 Weighted sum illustration . . . 26
1.14 E-constraint method . . . 27
1.15 Evolution Strategy . . . 29
1.16 CMA-ES algorithm . . . 32
1.17 CMA-ES algorithm (1+1) . . . 33
1.18 Crowding distance measure . . . 37
1.19 Kernel density estimation . . . 38

2.1 Overview of methods in computational chemistry . . . 42
2.2 Different layers from model to parametrization . . . 46

3.1 Master-worker scheme . . . 63
3.2 Water27 structures . . . 72
3.3 BEGDB water clusters . . . 73
3.4 Cluster population analysis . . . 74
3.5 Cluster population analysis . . . 75
3.6 QCE clusters . . . 75
3.7 Results from PM6 parametrization . . . 78
3.8 Total RMSD against energy error . . . 79
3.9 Solution vectors x_k after initial selection . . . 79
3.10 Final selection of parameter set . . . 80

3.11 Yang's γ-function 85
3.12 DFTB voids 86
3.13 Analytical γ-function 87
3.14 Results from DFTB3 optimization 88
3.15 Scatter-plot matrix for DFTB3 optimization 89
3.16 Total RMSD versus energy error 89
3.17 Radial distribution function for bulk water 93
3.18 Force-field densities 94
3.19 PM6 densities 95
3.20 DFTB3 densities 95
3.21 BLYP densities 96
3.22 Instantaneous Interface 99
3.23 Fourier transform of the interface fluctuation 99
3.24 Surface tension for the force-field methods from mechanical route 100
3.25 Surface tension for the semi-empirical approaches 101
3.26 Surface fluctuation for the classical approaches 102
3.27 Power spectrum of the DFTB3 semi-empirical method 103
3.28 Mean square displacement for the force-field methods 104
3.29 Mean square displacement for the semi-empirical approaches 104
3.30 Mean square displacement for the semi-empirical approaches 105

List of Tables

1.1 Default parameters for the (1+1)-CMA-ES 35

2.1 Definitions of the terms in the Hamiltonian 47

3.1 Task distribution to worker 64
3.2 Visualization of n-dimensional Pareto front 65
3.3 Demonstration of the standard algorithm for the k-mean clustering 66
3.4 Multi-objective set-up values for the most important parameters 68
3.5 Final reference set 75
3.6 PM6 parameters 76
3.7 Standard PM6 values for hydrogen 77
3.8 Standard PM6 values for oxygen 77
3.9 Standard PM6 diatomic values 77
3.10 The resulting candidates of the k-median clustering 81
3.11 Comparison of the final parameters with the original PM6 values 83
3.12 water density 94
3.13 Different density values obtained from NPT simulations 97
3.14 Surface tension from mechanical route 101
3.15 Different coarse graining for the instantaneous interface 102
3.16 Surface tension from instantaneous fluctuations 102
3.17 Different self-diffusion values from the NVE simulations of bulk water . . . 105

Introduction

Computational chemistry and molecular modeling are fast emerging areas which are used for the modeling and simulation of chemical and biological systems to understand and predict their behavior at the molecular level. Results can give insights in the structure of materials that are difficult or impossible to measure experimentally. Thus, computational chemistry has a wide range of applications in various disciplines [1] of engineering sciences, such as materials science, chemical engineering, bio-medical engineering, etc. Insights from molecular modeling is essential to understand the behavior of nano-systems [2]. It is probably the easiest route or gateway to the fast-growing discipline of nano-sciences and nanotechnology, which covers many areas of research dealing with manometer-sized objects and which is expected to revolutionize the industrial sector in the coming decades. Accurate chemical simulations often employ ab-initio methods where the expression ab-initio is the Latin term meaning "from the beginning". This name is given to computations which are derived directly from theoretical principles (such as the Schrödinger equation [3]), without inclusion of experimental data. Ab-initio methods, in fact, can be seen as an approximate quantum mechanical method. The approximations made are usually mathematical approximations, such as using a simpler functional form for a function, or getting an approximate solution to a differential equation. A common type of ab-initio calculation is called Hartree-Fock [4] calculation (HF), in which the primary approximation is called the mean-field approximation. In this approximation, the Coulombic electron-electron repulsion is approximated by a constant external electronic field caused by electrons. HF is a variational method, meaning that the calculated total energy of a molecule is equal to or greater than its exact energy. The second approximation in HF calculations is that the many electron wave-function must be described by some functional form, which is only known exactly for non-interacting electron systems [5]. The wave-function is formed from linear combinations of atomic orbitals, or more often from linear combinations of basis functions. Hence, for the linear combination and for the description of the orbitals, the use of numerical parameters is inevitable.
Embracing the use of parameter sets and reducing more functions to physical meaningful parameters results in the semi-empirical methods [6, 7, 8, 9]. Semi-empirical calculations are set up with the same general structure as a HF calculation. Within this framework, certain pieces of information, such as two electron integrals, are approximated or completely omitted. In order to correct for the errors introduced by omitting part of the

calculation, the method is parametrized, by curve fitting in a few parameters or numbers, in order to give the best possible agreement with experimental data or high level ab-initio calculations. The merit of semi-empirical calculations is that they are much faster than ab-initio calculations. The demerit of semi-empirical calculations is that the results can be slightly defective. If the molecule being computed is similar to molecules in the database used to parametrize the method, then the results may be very good. If the molecule being computed is significantly different from anything in the parametrization set, the answers may be very poor.

If a molecule is too big to effectively use a semi-empirical treatment, it is still possible to model its behavior by totally avoiding quantum mechanics. The methods, referred to as molecular mechanics [10], set up a simple algebraic expression for the total energy of a compound, with no necessity to compute a wave-function or total electron density. The energy expression consists of simple classical equations, such as the harmonic oscillator equation in order to describe the energy associated with bond stretching, bending, rotation, and inter-molecular forces, such as van der Waals interactions and hydrogen bonding [11]. All parameters in these equations must be obtained from experimental data or an ab-initio calculation.

The occurrence of parameters can not be avoided on any level of theory for a reasonable sized system. Although for the ab-initio methods the arising numerical parameters can be systematically derived, for the semi-empirical and classical methods this approach is unavailable. There is no standard method to determine and applied mathematics needs to be consulted.

Thus, in the present work, a re-optimization of a well established semi-empirical method is carried out with a threefold goal:

First, a systematic optimization approach is established. To date, all available parametrization of the semi-empirical methods is obtained by minimizing a constructed single function. This function is constructed by summing errors in energy, geometry, ionization and dipole amongst others. There is no clear consensus on how the function is build and also how the individual terms are weighted. This manuscript also points out some intrinsic defects when one combines objectives to one function. Because the present problem of parametrizing semi-empirical methods deals with various objectives to be optimized, the proper approach is to carry out multi-objective optimization. A detailed description of the multi-objective framework is given and the differences elaborated, that when carrying out multi-objective optimization not only a single solution but multiple trade-off solutions are found. In such a way, the subjective weighting of the objectives only enters the optimization process a posteriori and is completely detached from the optimization procedure itself.

The second goal is to demonstrate that bulk water properties can be reproduced from water cluster reference data (isolated system containing aggregated water molecules). Because the reference structures are crucial to the optimization process, a set from quantum cluster equilibrium theory [12] was used. With a reference set of 8 structures [13], already large improvements over standard values can be reported with the optimized

parameter set. To asses the performance of the optimized semi-empirical methods, we carried out some analysis of bulk and liquid-vapor water systems and compared them with the standard, ab-initio and classical methods in respect of radial distribution function, density, surface tension and mean square displacement.
The third goal is to reveal that the underlying theory of the semi-empirical model can capture all the necessary physics to describe bulk water systems. Thus, the electrostatic and dispersion terms are accurate enough to reproduce correct results and there is no need (as was carried out in the past) to correct for deficiencies because the resulting short comings are mainly based on incomplete parameter sets. Therefore it is crucial to explore more semi-empirical methods by investigating further the limits of the parameter space and not till then correcting for possible flaws in the underlying theory.
This thesis is organized in three chapters. A detailed and mathematical description of multi-objective optimization can be found in chapter 1. The theoretical framework of semi-empirical methods is elaborated in chapter 2 where the necessary physical and chemical properties are presented. In chapter 3, we apply the multi-objective optimization to re-parametrize the semi-empirical method. Various liquid water simulations are carried out with the newly obtained parameter set and compared to standard value calculations, ab-inito and classical force-field computations.

1 Optimization

This chapter reviews the basic theoretical concepts of parameter optimization. Starting from well known function optimization with common optimization techniques, the bridge is build to the case were the function possesses more than one objective. As the multi-objective optimization can not be seen as simple extension of the single-objective case, all the necessary concepts are outlined and presented to culminate in the review of algorithms to solve multi-objective problems. Hence, a short introduction to Evolution Strategy algorithms will be given and finally the Covariance Matrix Adaptation [14] for multi-objective optimization algorithm is presented.

1.1 Optimization Overview

The goal of optimization is to find one or more feasible solutions which correspond to extreme values of one or more objective functions or also just referred to as objectives [15]. In the simplest case, an optimization problem consists of maximizing or minimizing a real function by varying the input values from within an allowed set and computing the value of the function. The generalization of optimization theory to other formulations comprises a large area of applied mathematics [16]. However, optimization methods are of great importance in practice, particularly in computer science, engineering, economics, game theory and business decision-making [16]. But even the most fundamental principles in our world is the search for an optimal state. For that matter, the atoms in physics try to form bonds [17] in order to minimize the energy [17]. Consequently, the molecules form solid bodies during the process of freezing, trying to reach energy-optimal crystal structures.
In case of an optimization problem involving only one objective function, the task of finding the optimal solution is called *single-objective optimization* [16] and formulated as

follow:

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f(x) \\ \text{subject to} \quad & g_j(x) \leq 0, && j = 1,2,\ldots,J; \\ & h_k(x) = 0, && k = 1,2,\ldots,K; \\ & x_i^{(L)} \leq x_i \leq x_i^{(U)}, && i = 1,2,\ldots,n; \end{aligned}$$

whereas $f(x)$ is the *objective function* or simply *objective* to be minimized over the variable x. The equations $g_j(x) \leq 0$ are called inequality constraints and $h_k(x) = 0$ are termed equality constraints. Lower bounds $x_i^{(L)}$ and upper bounds $x_i^{(U)}$ on the variables x_i may exist and be treated explicitly to maintain feasibility.
Generally, one can distinguish between local [15] and global optimization problems whereas the latter goal is to find the global extrema of single-objective optimization problems. As example, a two-dimensional function $f(x_1, x_2)$ with one function value and thus one objective is shown in Figure 1.1 whereas local and global extrema exists.

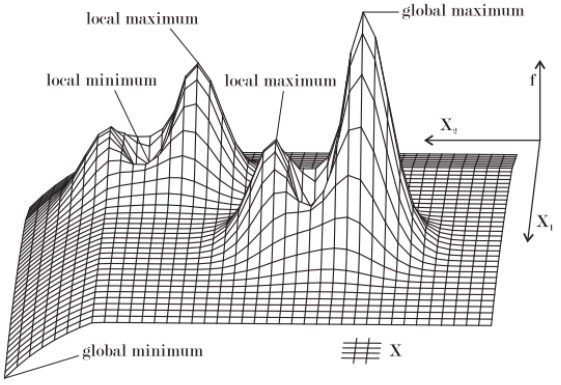


Figure 1.1 – Global and local optima of a two-dimensional function

In contrast to the global optimization problem, local search [18] is a meta-heuristic method for solving local optimization problems. The most prominent technique is called hill climbing [19] belonging to the family of local search. Examples of algorithms that solve convex (i.e convex functions) problems by hill-climbing include the simplex algorithm for linear programming and binary search. An other outstanding group of local optimization algorithms are represented by the gradient based methods [15]. Whereas the gradient descent (as illustrated in Figure 1.2), Newton's method [20], conjugated gradients [20] and the popular Broyden-Fletcher-Goldfarb-Shanno (BFGS) algorithm [20] are most known representatives. If the multi-variable function $F(x)$ is defined and differentiable in a neighborhood of a point a, then $F(x)$ decreases *fastest* if one goes from a in the direction of the negative gradient of F at a, $-\nabla F(a)$. It follows that, if $b = a - \gamma \nabla F(a)$ for γ small enough, then $F(a) \geq F(b)$. Consequently, one starts with a

guess x_0 for a local minimum of F, and considers the sequence $x_0, x_1, x_2, ...$ such that $x_{n+1} = x_n - \gamma_n \nabla F(x_n), n \geq 0$. and thus converges to the desired local minimum as explained in Figure 1.2.

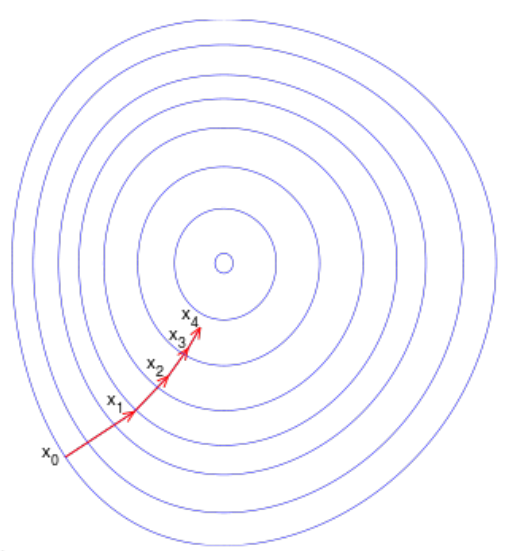


Figure 1.2 – Illustration of gradient descent.

Besides deterministic search principles involved in algorithms, there also exist stochastic search concepts [15]. These techniques are mostly used to find globally optimal solutions since using deterministically based algorithm would result in exhaustive enumeration of the search space and hence are not feasible even for relatively small problems. Most often the Monte Carlo algorithms [21] are applied which trade guaranteed correctness of the solution in for a shorter runtime. Thus, the results obtained with this technique are not totally incorrect - they may just not be the global optima in that respect. On the other hand, a slightly inferior solution is better than one which needs years to be found. Current trends in parallel processors [22] also advances the field of Monte Carlo algorithms since most of the problems are well suited to be solved in a massively parallel manner.

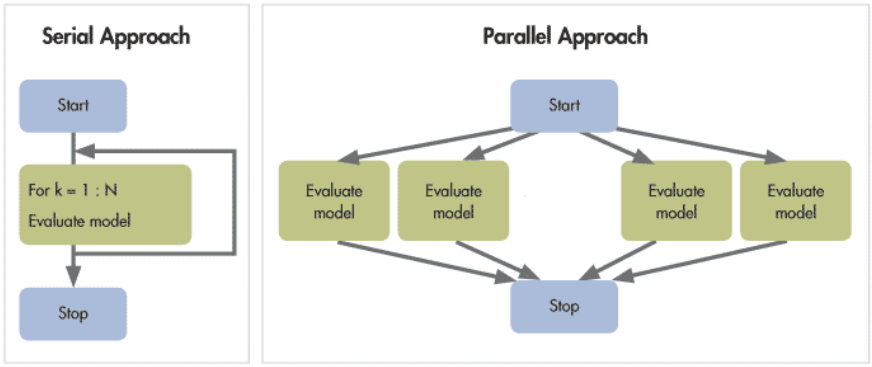


Figure 1.3 – Serial versus parallel approach

Figure 1.3 illustrates the embarrassingly parallel approach (little or no effort is required to separate the problem into a number of parallel tasks) for problems that can be solved in that fashion. The state of the art hardware for computation these days embrace multi-core and multi-processor computing arranged in clusters or grids (Figure 1.4), perfectly suited to solve parallel Monte Carlo problems. Hence, since all evolutionary algorithms (EA) are basic variation of the Monte Carlo algorithm (Figure 1.5), these application can take full advantage of the computing resources available and therefore opens up wide range of applications to be explored in the future.

Figure 1.4 – Piz Daint supercomputer [23]

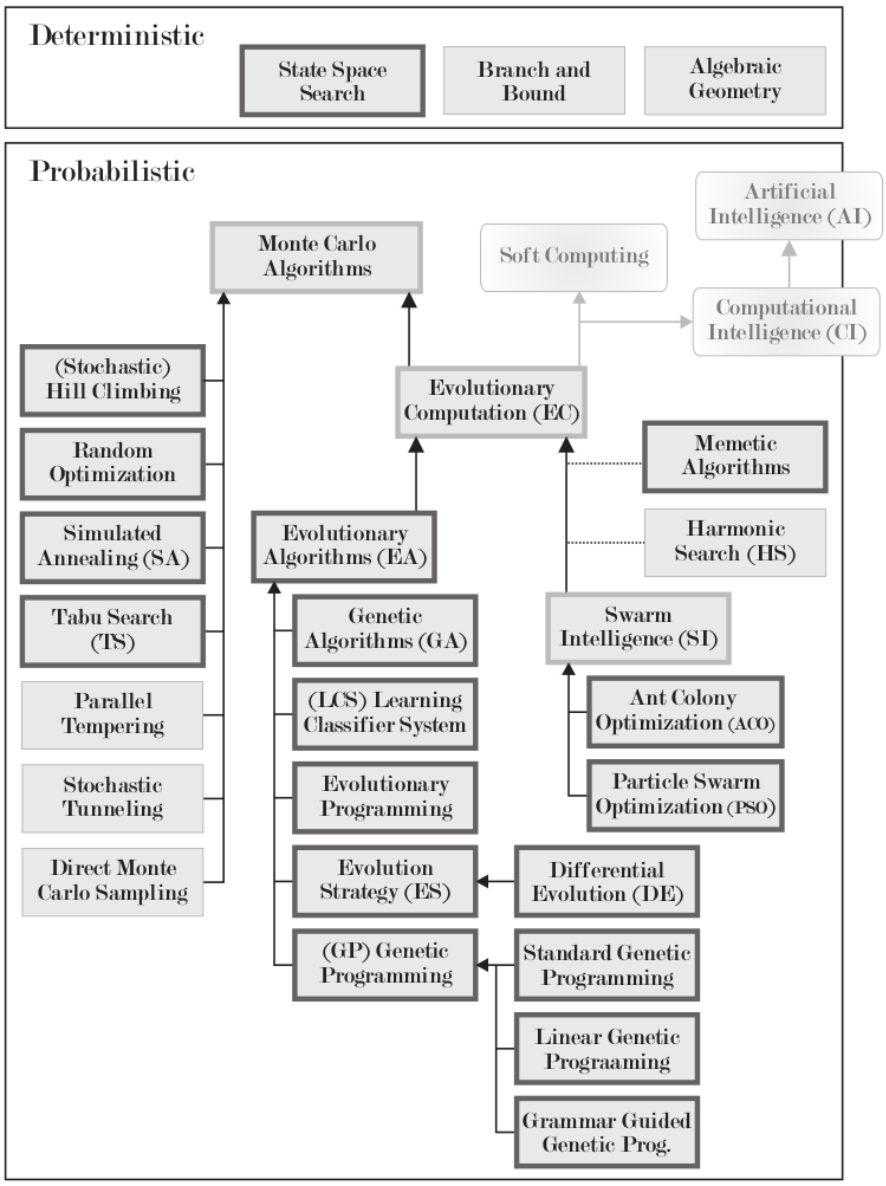


Figure 1.5 – Taxonomy of optimization algorithms

Most optimization algorithms use heuristics which are help functions to decide which one of a set of possible solutions is to be examined next. Heuristics can be applied by both, deterministic as well as probabilistic algorithms.
Meta-heuristics [15] usually work population-based or use a model of some physical process or natural phenomenon as heuristic function. An important class of Monte Carlo meta-heuristics is evolutionary computation [24]. It encompasses all such algorithms which are based on a set of multiple solution candidates (called population) which are iteratively refined. This field of optimization is also a group of soft computing [25] as well as a part of the artificial intelligence [26] area as pointed out in Figure 1.5. Its most

important members are evolutionary algorithms, which will be discussed in-depth in this chapter.
Beyond the evolutionary approaches, which are usually nature-inspired, there are also methods that mimic physical processes like simulated annealing and parallel tempering, as well as purely artificial techniques like tabu search [27] and random optimization.

1.2 Single and Multi-Objective Optimization

Most real-world search and optimization problems naturally involve multiple objectives [16]. Applying different sets to those objectives may produce trade-off's (conflicting scenarios) among different objectives. A solution that is extreme (in a better sense) with respect to one objective might mean a compromise in an other objective. This prohibits one to select a solution which is optimal with respect to only one objective.
Indifference curves in economics [28] or efficient frontiers in finance [29] are examples of real-world applications in the area of social science. There exists numerous problems in engineering such as optimal control [30] or various application to optimal design [31] to name a few. Since these problems are applied to sets F of n functions f_i, which represent multiple criteria [16] usually they are also referred to as multi-criteria problems.

$$F = \{f_i : X \to Y_i : 0 < i \leq n, Y_i \subseteq \mathbb{R}\} \tag{1.1}$$

Algorithms designed to optimize such a set F of objective functions are normally named with the prefix multi-objective, like multi-objective evolutionary algorithms [32].
Many algorithms and application case studies involving multiple objectives can be found. Unfortunately, most of those methods avoid the complexities involved in a true multi-objective optimization problem and transform multiple objectives into a single objective function by using some user-defined parameters. Hence, the majority of the studies do not treat multi-objective any different than single-objective optimization. Various studies concentrate on various means of converting multiple objectives into a single objective. This is contrary to the intuitive realization that single-objective optimization is a degenerate case of multi-objective optimization and clearly, multi-objective optimization can not be seen as simple extension of the single-objective case. However, there is a fundamental difference in single and multi-objective optimization which is neglected when using the transformation method. This important fact will be addressed in the following subsection.

1.2.1 Fundamental Differences

Since each objective corresponds to a different optimal solution, the result might be conflicting objectives. Lets consider a microchip manufacturer and the two main goals

are to decrease the size of the microprocessor and the overall costs. Thus, the production of smaller units intrinsically comes with higher costs because it is technically more demanding. This prohibits one to choose a solution which is optimal with respect to only one objective.

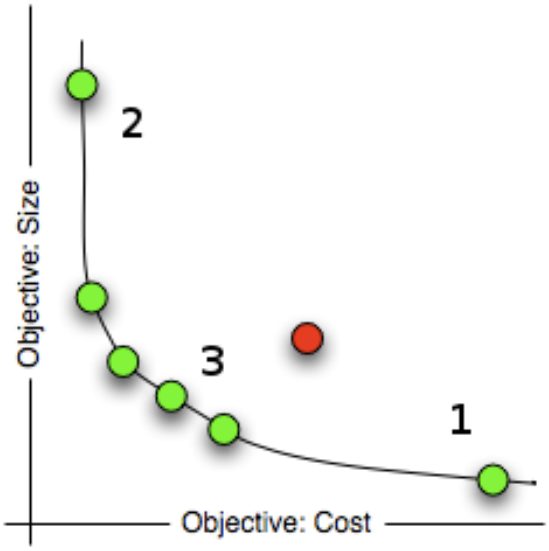


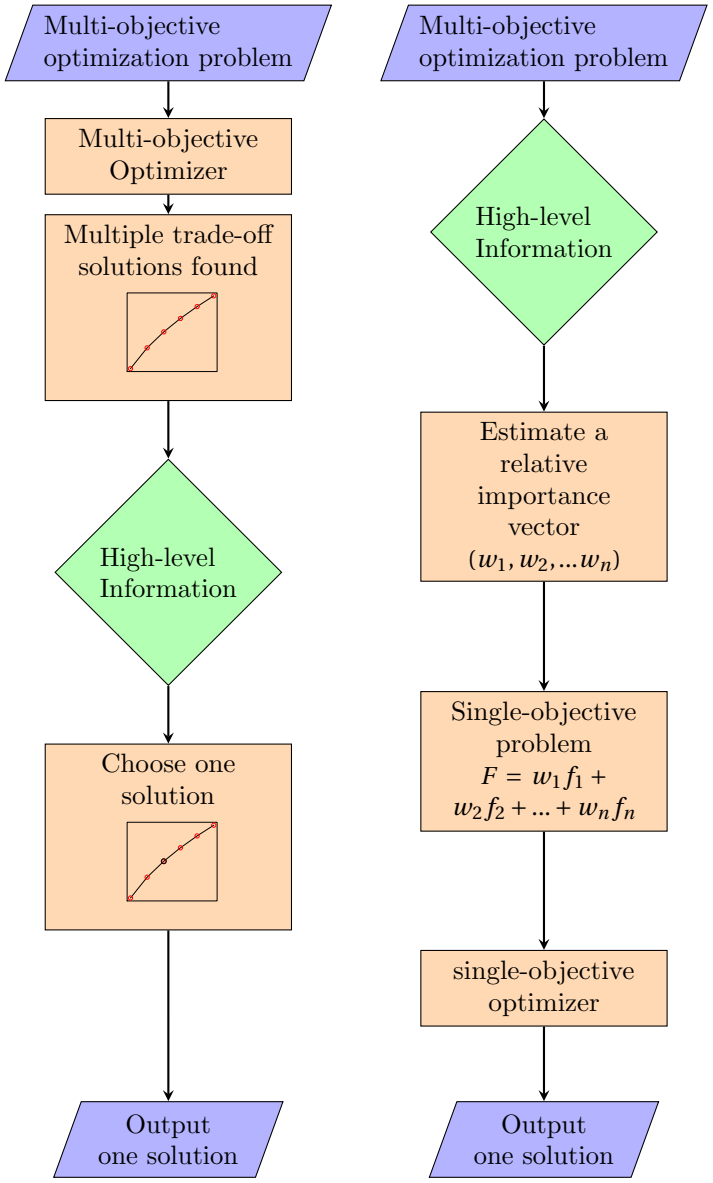
Figure 1.6 – Trade-off solutions for an multi-objective problem

In figure 1.6 if the only objective is the size, then the optimal solution would be point labelled 1. If cost would not matter, this chip type would be the only one considered. But given the economical element, the costs have to be included as well. This gives rise to multiple solution to the bi-objective problem of the manufacturer. Thus, between any two such solutions, one is better in terms of one objective, but this betterment comes only from a sacrifice on the other objective. We can state that none of these trade-off solutions is the best with respect to both objectives. Hence, in cases with more than one conflicting objective, there is no single optimum solution and without any further information, no solution from the set of optimal solutions can be said to be better than any other. This is the fundamental difference between a single-objective and a multi-objective optimization procedure.

1.2.2 Two Approaches to Multi-Objective Optimization

Although the underlying difference given the two optimizations lies in the cardinality in the optimal set, from a pragmatic standpoint, a user wants only one solution. Which of the resulting solutions must one choose? Getting back to the example from above, this is not an easy question to answer. For the microchip manufacturer it involves many other considerations, such as the total finances, costumer markets, research efforts, infrastructure, market position, and many other factors. Usually, this higher-level

information is of a non-technical origin and qualitative and experience-driven. Hence, in a multi-objective optimization, the effort must be made ideally in finding the set of trade-off optimal solutions followed by evaluation of higher-level consideration to make a choice.



(a) Schema of an ideal multi-objective optimization procedure

(b) Schematic of a preference-based multi-objective optimization procedure

Figure 1.7 – The two approaches to multi-objective optimization

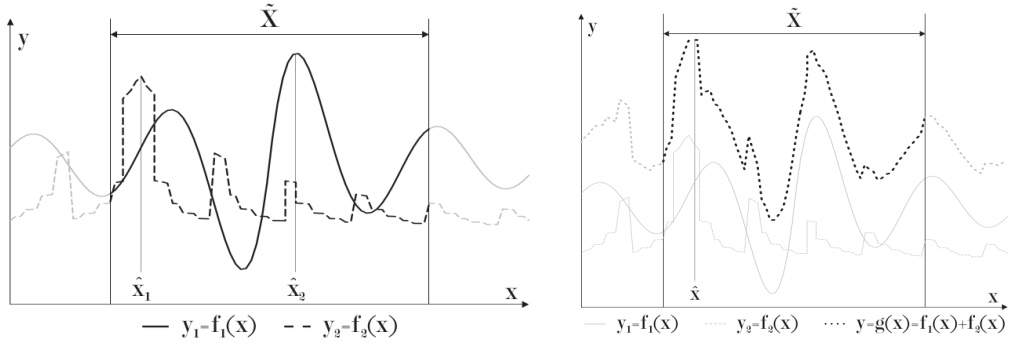
Figure 1.7 illustrates the principles of an ideal multi-objective process. In a first step, multiple trade-off solutions of the given multi-objective optimization problem are found.

Then, higher-level information are used to select a trade-off solution of the set. This corresponds to a specific order of importance of the objectives, and by selecting the solution at the end weighting them *a posteriori*. As can be seen in figure 1.6, solution point 1 assigns more importance to size than cost. On the other hand, solution 2 assigns more importance to cost than size. Hence, if such a relative preference factor among the objectives is known for a very specific problem, a simpler approach can be taken. This method would be to form a composite objective function as the weighted sum of the objectives, where the weight for an objective corresponds to the preference factor assigned to that particular objective. This scalarization approach assigns weights *a priori* to the multi-objective problem and results in a single-objective optimization problem which can easily solved with standard methods. A schematic of this *preference-based* multi-objective optimization is illustrated in figure 1.7. Based on higher-level information, the weight or preference vector w is constructed from the weights w_i representing the importance. Thereafter, the weight vector is used to construct the composite function, which is then used for the single-objective procedure.

$$g(x) = \sum_{i=1}^{n} w_i f_i(x) = \sum_{\forall f_i \in F} w_i f_i(x) \tag{1.2}$$

$$x^* \in X^* \leftrightarrow g(x^*) \geq g(x) \forall x \in \tilde{X} \tag{1.3}$$

It is clear that this method is extremely sensitive to the preference vector used for the weighted sum approach. Additionally, it is very difficult to find a relative preference vector since it is highly subjective and not straightforward. Without any knowledge of correlations or trade-off solutions, this is even a more difficult task. Yet another drawback of this approach is that it cannot handle functions that rise or fall with different speed [16] properly. For instance, the sum of $f_1(x) = -x^2$ and $f_2(x) = e^{x-2}$ will always disregard one of two functions, depending on the interval chosen. For small x, f_2 is negligible compared to f_1. On the other hand, for $x > 6$ it begins to outpace f_1 which, in turn, will become unimportant. It is not possible to add up properly such functions using constant weights. Even setting w_1 to a very large number such as 10^{10} will fail because f_1 will become insignificant for $x > 40$, because $\left|\frac{-40^2 \times 10^{10}}{e^{40-2}}\right| \sim 0.0005$. Therefore, preference-based approaches are only suitable to optimize functions that share the same big $\mathcal{O}$ notation. A similar problem is the scalarization of different units. In practical cases, the different objectives f_i consist of different physical units. To construct the composite function one needs to normalize these functions first and it is not evident how to proceed and rather what normalization constant should be used.
Figure 1.8b illustrates the optimization using the preference-based or weighted sum approach for the example in figure 1.8a. The weights are taken as 1, which maximizes both functions f_1 and f_2 and leads to a single optimum $x^* = \tilde{x}$.



(a) Two functions f_1 and f_2 with different maxima $\tilde{x_1}$ and $\tilde{x_2}$

(b) Optimization using the weighted sum approach with final single-objective function

Figure 1.8 – Illustration of the weighted sum approach

Historically, a lot of time and energy was devoted to single-objective optimization [16] and therefore it is not surprising that the preference-based advances in multi-objective optimization gained a lot of popularity [16]. That is, especially the weighted sum approach was extensively used also in quantum chemistry [33] to parametrize semi-empirical methods. Additionally, the weighted sum approach was motivated by the fact that this method would find only one single solution in a simulation run and therefore simplify the complexity of the given problem. Of course this simplification is a fallacy, as the higher-level information to construct the composite function is never complete and hence a vast number of artefacts are introduced. But since the method produced reasonable and usable results [33], it can be a suitable approach. Nevertheless, the lack of a systematics and methodology makes it almost impossible to verify and reproduce preference-based optimization procedure reported in literature. Therefore, real multi-objective optimization is advised where the entire process is transparent and even when the high-level information are changing the optimization process does not have to be repeated.
Indeed, the area of optimization has changed rapidly over the last few years due to stochastic techniques [15]. Both, the field of global optimization as well as the multi-objective optimization domain was influenced massively by the *evolutionary algorithms* (EA). The advantage of those methods, which will be disused extensively in this manuscript, is that a population of solutions is maintained and hence as a result a set of solutions is produced. Ideally, it can be expected that the members of the population can cover the set of optimum solutions. Accordingly, the goal of multi-objective evolutionary algorithms is to find a diverse set of optimum solutions in its final population. Since for solving multi-objective optimization problem we need different sets of optimal solutions, this makes the evolutionary algorithms extremely suitable to this kind of problem setting.

1.3 Multi-Objective Optimization

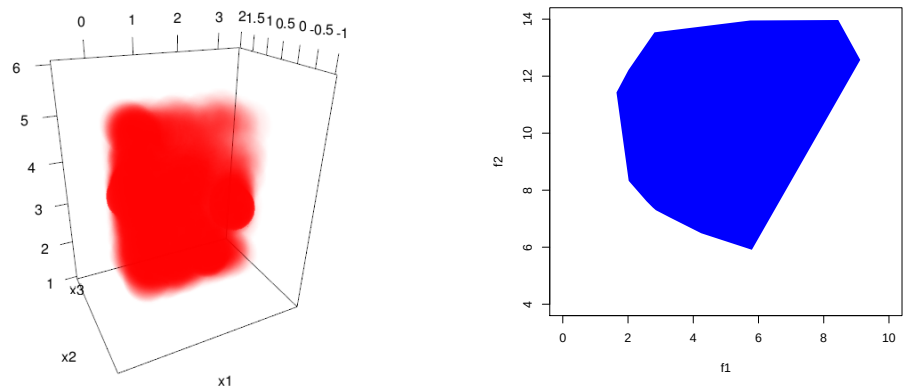
A multi-objective optimization problem deals with multiple criteria decision making, that is concerned with mathematical optimization problems involving more than one objective function to be optimized simultaneously [16]. Historically, those problems were mostly cast into and solved as a single-objective optimization problem due to lack of suitable solution methodologies. However, this conversion introduces a huge amount of artefacts and is extremely error prone. Moreover, multi-objective optimization can not be seen as simple extension of the single-optimization case. Since in the single-optimization case only one solution is obtained, it is wrong to assume that for the multi-objective case there would be one solution corresponding to each objective function. In the following section, we will discuss the fundamental principles of multi-objective optimization.

1.3.1 Basic Concepts

A multi-objective optimization problem is an optimization problem that involves multiple objective functions [16]. In mathematical terms, a multi-objective optimization problem can be formulated as

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & f_m(x), && m = 2,3,\ldots,M; \\ \text{subject to} \quad & g_j(x) \geq 0, && j = 1,2,\ldots,J; \\ & h_k(x) = 0, && k = 1,2,\ldots,K; \\ & x_i^{(L)} \leq x_i \leq x_i^{(U)}, && i = 1,2,\ldots,n; \end{aligned}$$

where the integer $m \geq 2$ is the number of objectives, and the minimization is subject to J inequality constrains as well as K equality constrains. The last set of constraints are called variable bounds, restricting each variable x_i to be within boundaries $x_i^{(L)}$ and $x_i^{(U)}$. All solutions fulfilling the constraint functions are called feasible set. Hence an element $x^* \in S$ is called a feasible solution or a feasible decision [16] where S is the feasible decision space $S \subset \mathbb{R}^n$. Therefore, we realize that in the presence of constraints, the entire decision variable space D does not need to be feasible. In general, a solution x is a vector of n decision variables: $x = (x_1, x_2, \ldots, x_n)^T$. In addition, the vector-valued objective function is often defined as $f : S \to \mathbb{R}^m, f(x) = (f_1(x), \ldots, f_k(x))^T$. From this formulation it is more evident why multi-objective optimization is also referred to as vector optimization or multi-attribute optimization. This is also the most eminent feature of multi-objective optimization where there is a mapping from $f : \mathbb{R}^n \to \mathbb{R}^m$. In contrast the single-objective case where the mapping is $f : \mathbb{R}^n \to \mathbb{R}$. Figure 1.9 illustrates an example of a the mapping from the decision space ($S \subset \mathbb{R}^3$) with the variables x_1, x_2, x_3 to the objective space with the two objectives f_1, f_2.



(a) Decision space with the variables $x = (x_1, x_2, x_3)$

(b) Objective space with the two objectives $f = \{(f_1, f_2) \in D\}$

Figure 1.9 – Mapping from decision space to objective space

1.3.2 Difference with Single-Objective Optimization

Besides the obvious difference of having multiple objectives, there are a number of fundamental differences between single-objective and the multiple-objective counterpart as follows:

- two goals instead of one;
- dealing with different search spaces;
- no artifical fix-ups;

We will take a look at those differences in more detail in the following subsection.

1.3.2.1 Two Goals Instead of One

As we have seen, in multi-objective optimization, there are clearly two goals compared to only finding the best solution in single-objective cases. The first aim is to progress towards the front of the best trade-off solutions [34, 35], which is certainly an important goal. Nevertheless, it is equally important to maintain a diverse set of solutions in the non-dominated front since an algorithm that finds a closely packed set of solutions would not be very useful. Given that *all* objectives are important in a multi-objective optimization, manifold solutions close to the Pareto-optimal front provide different sets that deal with the different trade-off for the given objectives. These two goals, progressing towards the Pareto-optimal front and maintaining diversity, are somewhat *orthogonal* to each other. Therefore mechanisms to converge to the Pareto-optimal front as well

as maintenance of a diverse set of solutions must be incorporated in an algorithm to solve multi-objective problems as desired. On account of this dual tasks, multi-objective optimizations are much harder to solve than the single-objective problems.

1.3.2.2 Dealing with Different Search Spaces

Yet another hurdle is that a multi-objective optimization problem has to deal with different search spaces, instead of one. In single-objective optimization, there is only one search space and the solutions are either accepted or rejected according to the objective function values. Clearly, $f : \mathbb{R}^n \rightarrow \mathbb{R}$ as it is present in single-objective optimization is a very simple one-dimensional criterion space $\mathbb{R}$. On the other hand the criterion space in the multi-objective optimization is fare more complex ($\mathbb{R}^m$) and the requirement on an algorithm is that the proceedings in all spaces must be coordinated in such a way that the creation of new solutions in the decision variable space is complimentary to the diversity need in the objective space. Obviously, this is a very difficult task and depends very much on the mapping between the spaces.

1.3.2.3 No Artificial Fix-Ups

As we have already briefly mentioned, historically owing to the fact of lacking suitable means of handling multi-objective optimization problems, different fix-ups were invented. The most prominent approach, the weighted sum proposal, introduces an artificial fix-up through the chosen weights. Multi-objective optimization for finding multiple Pareto-optimal solutions on the other hand, eliminates all such fix-ups. Moreover, especially in the field of research, knowledge of such multiple optimal solutions may help a designer to compare solutions and getting an understanding of the underlying problem. Since there is no need to conduct multiple runs and also no artificial fix-ups are introduced, and above all, the concept of dominance (which will be introduced shortly) can be used, all this helps to overcome some of the difficulties and introduces a powerful tool to handle multiple objectives.

1.3.3 Dominance and Pareto Optimality

The majority of multi-objective optimization algorithms use the concept of dominance for ranking the solutions. In this subsection we investigate the concept of dominance and related terms and present some techniques for finding dominant results in a finite set of solutions.

1.3.3.1 Concept of Domination

We use the following paragraph to describe the concept of domination since this is usually needed to compare two solutions.
The operators $\lhd$ and $\rhd$ are used to denote if an solution i is better or worse than a solution j on a particular objective. Therefore $i \lhd j$ indicates that the solution i is better than j for a given objective.

Definition 1. A solution $x^{(1)}$ dominates an other solution $x^{(2)}$ if follwoing conditions are fullfiled:

- The solution $x^{(1)}$ is no worse than $x^{(2)}$ in all objectives, or if $f_j(x^{(1)}) \not\rhd f_j(x^{(2)})\ \forall j = 1,2,\ldots,M$
- The solution $x^{(1)}$ is strictly better than $x^{(2)}$ in at least one objective, or $f_j(x^{(1)}) \lhd f_j(x^{(2)})\ \exists\ j \in \{1,2,\ldots,M\}$.

In case that both conditions are satisfied we can conclude that solution $x^{(1)}$ dominates the solution $x^{(2)}$. Mathematically this is equivalent to stating $x^{(1)} \preceq x^{(2)}$ and hence $x^{(2)}$ is dominated by $x^{(1)}$.

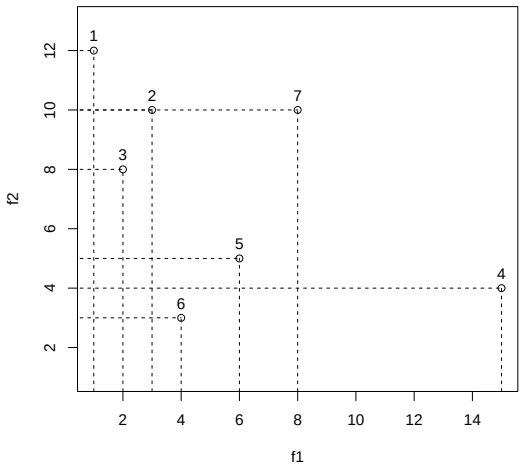


Figure 1.10 – A population of seven solutions

We shall consider a bi-objective optimization problem with a given set of solutions as shown in Figure 1.10. Since both functions f_1 and f_2 need to be minimized, it is difficult to decide which is the best solution in the set. However, applying the concept of domination to decide which solutions are better can clarify the situation. One can for example see that solution 6 is dominating solutions 4,5 and 7. We can also observe that solution 3 is

better in terms of both objectives than solution 2 and 7. Solutions 2 and 7 are equally good in objective f_2 but since solutions 2 is better in respect to the objective f_1 it satisfies the definition of dominance and hence solution 2 dominates solution 7.
It is important for an optimization algorithm to keep the best set of solution, thus the concept of domination offers a tool for this comparison and finally provides a selection of non-dominated candidates. In the case of the example in Figure 1.10, the solutions $\{1,3,6\}$ are superior and therefore should be selected over the other solutions $\{2,4,5,7\}$. Most of the multi-objective optimization algorithms use this non-domination principle as concept for selection as we shall see in this chapter.

1.3.3.2 Concept of Pareto-Optimality

Performing all pair-wise comparisons and noting which solution dominates others and which solutions are non-dominated with respect to each other for a given set of solutions results in a set of solutions which do not dominate each other. This particular set of solution is usually termed *non dominated set* for a given solution space [16]. In the example of Figure 1.10 the non dominated set is represented as $\{1,3,6\}$ since non of these solutions are dominated by any other solution in the set. Thus, a set of non-dominated solutions is defined as follows [16].

Definition 2. Among a set of solutions P, the non-dominated set of solutions P' are those that are not dominated by any member of the set P

In case the set P is the entire search space ($P = S$), we call the resulting non-dominated set P' the *Pareto-optimal set.* The term is named after Vilfredo Pareto (1848-1923) [36], an Italian economist who used the concept in his studies of economic efficiency and income distribution.
We illustrate an bi-objective optimization problem with objectives f_1 and f_2 respectively which both should be minimized. Analysing all possible pair-wise comparisons of the solutions $x_i \overset{?}{\lhd} x_j, i \neq j$ results in a set of non dominated solutions $\{15,18,22,24,46\}$ as shown in 1.11. The solid red lines in Figure 1.11 mark the *Pareto-front* that is the connection of the set of non-dominated solutions.

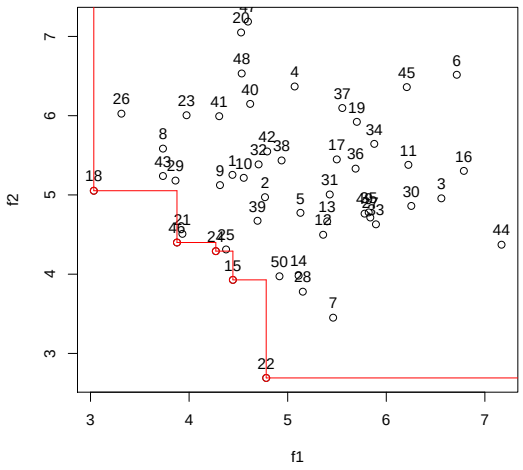


Figure 1.11 – An illustration of the Pareto-front of the non dominated set of solutions

Since there can be local Pareto-optimal sets in multi-objective optimization similar to local optimal solutions in the case of single-objective optimization, we define the global Pareto-optimal set as accordingly [16].

Definition 3. The non-dominated set of the entire feasible search space S is the globally Pareto-optimal set.

In the literature, usually the globally Pareto-optimal set is simply referred to as the Pareto set [16] because within the feasible search space those candidates are not dominated by any member of the set and hence are the optimal solutions to the given multi-objective optimization problem.

1.3.3.3 Algorithms for Finding a Non-Dominated Set

A very slow and naive approach for finding a non-dominated set could be to compare each solution x_i with every other solution x_j in the population to verify if it is dominated. If a solution x_j is found to dominate solution x_i, it is clear that it can not belong to the non-dominated set and hence it is flagged and one advances to the next iteration i. The following approach describes an algorithm for finding a non-dominated set in a given set P of size N.

Algorithm 1: Naive and slow approach

Data: set P
Result: non-dominated set P'

```
i=1;
N=size(P);
P'={ };
while i ≤ N do
    x_i = P_i;
    while j ≤ N; i ≠ j do
        x_j = P_j;
        if x_j ◁ x_i then
            dominated=true;
            exit;
        else
            dominated=false;
        j = j + 1
    if (not dominated) then
        P' = P' ∪ {x_i}
    i = i +1
```

We illustrate the scheme of the algorithm on the same set of seven ($N = 7$) solutions, as shown in Figure 1.10. The result of the algorithm is the non-dominated set of solutions P'. As seen previously, this should include the solutions $\{1,3,6\}$ and indeed this is exactly the result we obtain applying the working principle presented. Following a step by step illustration is given for the simple and naive approach.

Step 1: $i = 1; P' = \{\}$

Step 2: x_1 is compared to all other solutions $x_j; i \neq j$ of the set

Step 3: since non of the solutions x_j dominate x_1, we add $x_1 \cup P'$

Step 4: x_2 is compared to all other solutions $x_j; i \neq j$ of the set

Step 5: it is found that solution x_3 dominates x_2 and hence x_2 is flagged dominated and we skip to the next solution

Step 6: x_3 is compared to all other solutions $x_j; i \neq j$ of the set

Step 7: since non of the solutions x_j dominate x_3, we add $x_3 \cup P'$

Step 8: x_4 is compared to all other solutions $x_j; i \neq j$ of the set

Step 9: it is found that solution x_6 dominates x_4 and hence x_4 is flagged dominated and we skip to the next solution

Step 10: x_5 is compared to all other solutions $x_j; i \neq j$ of the set

Step 11: it is found that solution x_6 dominates x_5 and hence x_5 is flagged dominated and we skip to the next solution

Step 12: x_6 is compared to all other solutions $x_j; i \neq j$ of the set

Step 13: since non of the solutions x_j dominate x_6, we add $x_6 \cup P'$

Step 14: x_7 is compared to all other solutions $x_j; i \neq j$ of the set

Step 15: it is found that solution x_2 dominates x_7 and hence x_7 is flagged dominated and we stop since $i = 7 = N$

As expected, the resulting non-dominated set is $P' = \{1, 3, 6\}$.
The inner loop results in $\mathcal{O}(N)$ comparisons for domination and each comparison consists of M function value comparisons. Therefore, the total complexity of the inner loop is $\mathcal{O}(MN)$. Recursively, the outer loop adds additional $\mathcal{O}(N)$ operations, resulting in a total worst case complexity of $\mathcal{O}(MN^2)$. Since in the inner loop most likely not all $N-1$ solutions have to be checked before a dominant solution can be found, the real complexity might be smaller than $\mathcal{O}(MN^2)$.
Although the worst case complexity of all algorithms for finding non dominated set of a multi-objective optimization problem is $\mathcal{O}(MN^2)$, there exists a variety of more efficient algorithms [2], for one the method proposed by Fang et el [37] and more recently refined algorithms by Ding [38] and Jun Du [39] and Mishra [40]. The best case time complexity of the latter is $\mathcal{O}(Nlog(N))$ and hence an improvement as compared to other algorithms.

As we shall see later, some multi-objective optimization algorithms require us not only to find the best non-dominated front in a population, but to be classified into various non-domination levels. Hence, the population needs to be sorted according to an ascending level of non-dominance.

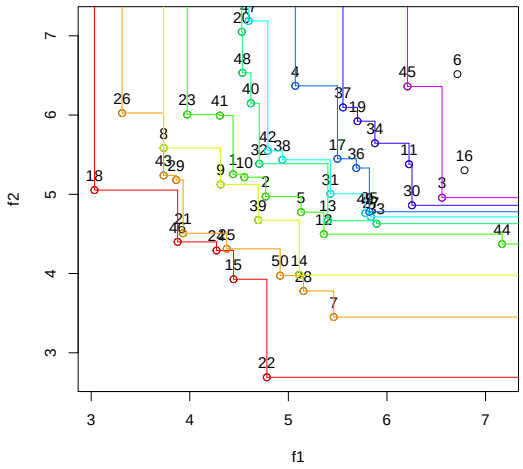


Figure 1.12 – Different pareto-levels for a set of 50 solutions of a multi-objective optimization problem

Usually, the best solution of non dominance are called non-dominant solutions of level 1. Removing this set from all the solution and repeating the procedure gives solution to the non-dominant set of level 2. This process can be repeated until all solutions are assigned a level of dominance. This is equivalent with designating each solutions to a corresponding Pareto-front of different levels. Figure 1.12 illustrates the same set of solution as Figure 1.11 but with corresponding non-dominance level ranging from 1 to 9 (from red to purple). The different Pareto fronts are recursively produced removing the leading Pareto-front and repeating the procedure to find the remaining non-dominant set of the population.

As before, the procedure requires $\mathcal{O}(MN^2)$ computations per iteration and since it is repeated to find the subsequent fronts, the worst case (when there exists only one solution in each front) complexity of the non-dominant sorting algorithm is $\mathcal{O}(MN^3)$. However, Deb et al [16] described a fast non-dominated sorting approach which will require at most $\mathcal{O}(MN^2)$ computations.

For each solution two entities are calculated: (i) n_i, the number of solutions which dominate the solution i, and (ii) S_i, the set of solutions which are dominated by solution x_i. To calculate these entities $\mathcal{O}(MN^2)$ comparisons are required. All points with $n_i = 0$ are put in a list F_1 which is called the current or pivoting front. Subsequently, for each solution in the current front we visit each member (j) in the sets S_i and reduce its n_j

count by once since the solution is not in the set any more. Due to this, if for any member j the count n_j becomes zero, we put it in a separate list H. After checking all members of the current front, we declare F_1 as the first front. Hence, the process is continued using the newly identified front H as the current front.
This iteration requires $\mathcal{O}(N)$ computations and this process is carried out till all fronts are identified. In a worst case there can be N fronts and consequently the worst case complexity of this loop is $\mathcal{O}(N^2)$ resulting in a overall complexity of the algorithm as $\mathcal{O}(MN^2)+\mathcal{O}(N^2)$ or $\mathcal{O}(MN^2)$.
Although the computational burden was reduced from $\mathcal{O}(MN^3)$ to $\mathcal{O}(MN^2)$ by performing systematic book-keeping, the storage requirements were increased from $\mathcal{O}(N)$ to $\mathcal{O}(N^2)$ in the worst case scenario. The fast non-dominated sorting algorithm applied on a population P returning a list of non-dominated fronts F is outlined as following.

Algorithm 2: fast non-dominated-sort (P)

Data: set P
Result: lists of non-dominated sets L_i
i=1;N=size(P);L={ };
while $i \leq N$ **do**
  $x_i = P_i$;
  while $j \leq N; i \neq j$ **do**
    $x_j = P_j$;
    if $x_j \lhd x_i$ **then** ; // if x_j dominates x_i then
      $S_{x_i} = S_{x_i} \cup x_j$; // include x_j in S_{x_i}
    else if $x_i \lhd x_j$ **then**; // if x_i is dominated by x_j then
      $n_{x_i} = n_{x_i} + 1$; // increment n_{x_i}
    j = j + 1
  if $n_{x_i} = 0$ **then**
    $F_1 = F_1\{\cup x_i\ \}$
  i = i +1
while $F_i \neq \{\ \}$ **do**
  $H = \{\ \}$;
  N=size(F_i);
  while $i \leq N$ **do**
    K=size(S_{x_i});
    while $j \leq K$ **do**
      $n_{x_j} = n_{x_j} - 1$;
      if $n_{x_j} = 0$ **then**
        $H = H \cup x_j$
      ; // if n_{x_j} is zero, x_j is a member of list H

1.3.4 Classical Methods

In this subsection we will describe common classical approaches used in multi-objective optimization. Thus, because these methods are in contrast to the *a posteriori* methods such as evolutionary algorithms and hence follow the path given in Figure 1.7b, one refers to this group as classical methods. Classical approaches have been around for at least the past four decades and many different algorithms have been suggested [16]. Cohon [41] or Hwang and Masud [42] and later Miettinen [43] have attempted to classify the algorithms to various considerations. Here we introduce only the following three classes:

- Scalarizing methods
- No-preference methods
- A priori methods

We will illustrate the first class with the most prominent example. It is important to note that none of the classical approaches make any attempt to find multiple Pareto-optimal solutions. Hence, the classical methods aim at finding a single optimal solution which is ill-defined for a multi-objective optimization problem.

1.3.4.1 Scalarizing multi-objective Optimization Problems

We already introduced the weighted sum method which is a particular case of the more general scalarizing approach to multi-objective optimization. The more general formulation for a scalarization for a multi-objective optimization is thus

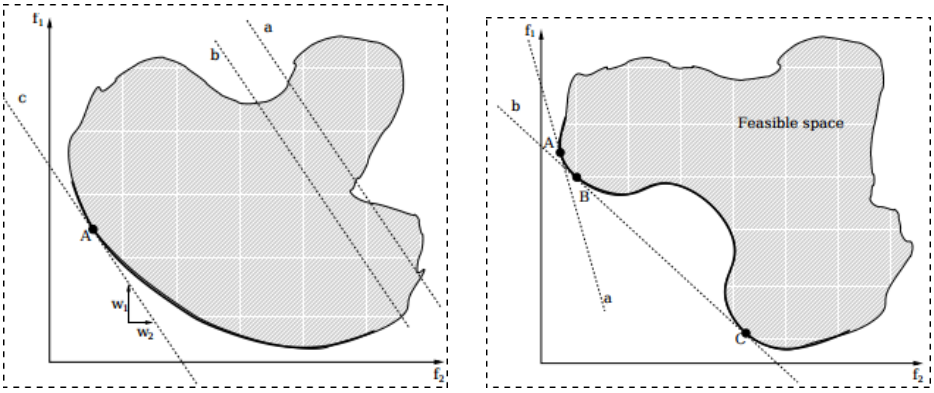
$$\begin{aligned} \underset{x}{\text{minimize}} \quad & F(f_m(x)), && m = 1,2,\ldots,M; \\ \text{subject to} \quad & g_j(x) \geq 0, && j = 1,2,\ldots,J; \\ & h_k(x) = 0, && k = 1,2,\ldots,K; \\ & x_i^{(L)} \leq x_i \leq x_i^{(U)}, && i = 1,2,\ldots,n; \end{aligned}$$

where F is the scalarization function $F : \mathbb{R}^m \to \mathbb{R}$. The simplest approach and probably also the most widely used scalarization function is the **weighted sum or linear scalarization method** where $F = \sum_{m=1}^{M} w_m f_m(x)$. As the name suggests, each objective gets pre-multiplied by a user-supplied weight-vector w_m. As simple and elegant this approach may seem, it introduces the difficult question of what weights must one use. The answer depends on the importance of each objective as well as the scaling factor or normalization (i.e to render unit-less quantities). The *normalization* of each objective is far from trivial but crucial for the weighted sum approach. Ideally, each objective should be scaled in a way rendering all ranges appropriate and similar. Hence, the linear scalarization can be

written as

$$\begin{aligned} \underset{x}{\text{minimize}} \quad & F(x) = \sum_{m=1}^{M} \frac{w_m}{n_m} f_m(x), && m = 1,2,\ldots,M; \\ \text{subject to} \quad & g_j(x) \geq 0, && j = 1,2,\ldots,J; \\ & h_k(x) = 0, && k = 1,2,\ldots,K; \\ & x_i^{(L)} \leq x_i \leq x_i^{(U)}, && i = 1,2,\ldots,n; \end{aligned}$$

where $w_m \in [0,1]$ are the weights or importance of each objective (usually the additional constraint $\sum_{m=1}^{M} w_m = 1$ is applied) and n_m is the normalization such that the objectives are represented in the same order of magnitude. A number of interesting theorems regarding the relationship between the optimal solutions of the linear scalarization and the true Pareto-optimal solutions in classical texts exits [16]. We only note that the weighted sum approach represents one Pareto-optimal solution if the weight is positive for all objectives. The same is true for the reciprocal way namely, if x^* is a Pareto-optimal solution of a convex problem, then there exits a non-zero positive weight vector w such that x^* is a solutions to the problem. The proof can be found in Miettinen [43] and as consequence for a convex multi-objective optimization problem any Pareto-optimal solutions can be found by using the weighted sum method as illustrated in 1.13a.



(a) Geometrical representation of the weight-sum approach in the convex Pareto curve case

(b) Geometrical representation of the weight-sum approach in the non-convex Pareto curve case

Figure 1.13 – Different geometrical shapes of the Pareto curve

The above linear scalarization approach cannot find certain Pareto-optimal solutions in the case of a non-convex objective space. Figure 1.13b explains this scenario whereas no weight can be chosen such that a contour line results in the region between B and C. This is because before a line becomes a tangent to any point in BC, it gets a tangent elsewhere and hence the Pareto-front in the region BC can not be found by the weighted sum method. Unfortunately this property has been neglected in many studies because an convex objective space was assumed which of course is not generally applicable and

therefore caution must be applied when using the linear scalarization method.
An alternative scalarization method offers a solution to overcome the problem of non-convex Pareto optimality. The **ϵ-constraint method** as suggested by Haimes et al [44] reformulates the multi-objective optimization problem by just keeping one of the objectives and restricting the rest of the objectives within predefined and user-specified values. The modified problem is as follows:

$$
\begin{aligned}
\underset{x}{\text{minimize}} \quad & f_\mu(x), \\
\text{subject to} \quad & f_m(x) \le \epsilon_m \qquad m = 1,2,\ldots,M; \\
& g_j(x) \ge 0, \qquad j = 1,2,\ldots,J; \\
& h_k(x) = 0, \qquad k = 1,2,\ldots,K; \\
& x_i^{(L)} \le x_i \le x_i^{(U)}, \quad i = 1,2,\ldots,n;
\end{aligned}
$$

where the parameter ϵ_m corresponds to an upper bound of the value of f_m and does not necessarily have to be a value close to zero. As illustration in figure 1.14 let us say that we retain f_1 as an objective and treat f_2 as a constraint: $f_2 \le \epsilon$ with different values for ϵ. Let us consider the scenario with $\epsilon = \epsilon_4$ in figure 1.14. The left portion becomes the feasible solution to the original bi-objective optimization problem. It can be seen that the minimum solution to this constraint problem would be solution 'C'. For $\epsilon = \epsilon_3$ the resulting minimum solution would be found in point 'B' and consequently for $\epsilon = \epsilon_2$ the best solution would be 'A'.

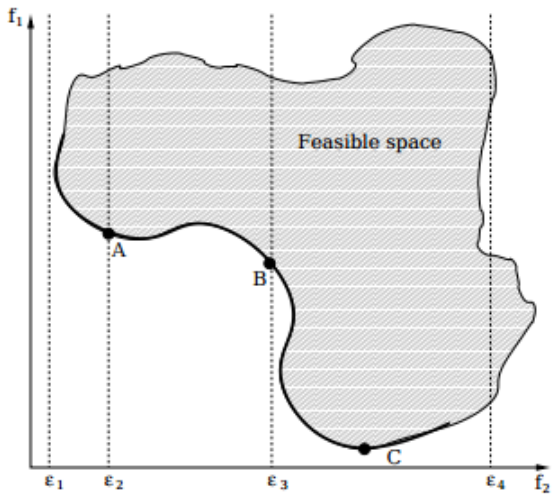


Figure 1.14 – A non-convex Pareto front with different ϵ values and corresponding solution points.

This approach is able to identify a number of non-inferior solutions on a non-convex boundary that are not obtainable using the weighted sum technique. A disadvantage of this approach is that the use of hard constraints is rarely adequate for expressing

true design objectives. The optimization proceeds with reference to these priorities and allowable bounds of acceptance. The difficulty here is in expressing such information at early stages of the optimization cycle.
For other scalarization approaches such as the Guddat [45] or Benson [46] method, the min-max scalarization [47], Chebyshev-norm approach as well as Pascelotti-Serafini scalarization [48] we refer to [16] for a detailed description. For an additional summary of the no-preference and a priori methods the reader is advised to consult [16] or [49].

1.3.4.2 Summary and Disadvantages of the Classical Approaches

Some major difficulties arise from the classical approaches as we have seen in the above section. Most algorithms convert the multi-objective optimization problem into a single-objective optimization problem by using some user-defined methods. Of these methods usually employed, the weighted sum approach converts the objectives into a single objective by normalizing and weighting and summing each objective. The preference vector requires some problem knowledge to impose an a priori weighting. It can be shown that the linear scalarization method is incapable of finding trade-off solutions in problems with non-convex Pareto-fronts [16]. Although the ϵ-constraint method overcomes this pitfalls, the ϵ-vector to be selected highly influences the outcome and has to be dealt with great elaborateness. Therefore, mainly three problems remain:

- Only one Pareto-optimal solutions can be found in one simulation run
- Some Pareto-optimal solutions can not be found in non-convex multi-objective optimization problems
- All classical algorithm require some *a priori* knowledge of the problem

1.3.5 Evolutionary Algorithm and Covariance Matrix Adaptation

Evolutionary algorithms (EA) are based on adapting Darwinian principles and therefore mimic natural evolution to constitute search and optimization procedures, hence the name. As we have seen in Figure 1.7 technically they belong to the family of Monte Carlo (Figure 1.5) or trial and error problem solver in the class of probabilistic methods. Nevertheless, the are some differences from classical search and optimization procedures in a variety of ways for the evolutionary algorithms. The focus in this chapter will be on the special class of Evolution Strategy (ES) which is a subgroup of the more general evolutionary algorithms as shown in Figure 1.5.
Most multi-objective optimization algorithms are based on single-objective algorithms in special ways, thus an understanding of the single-objective evolutionary algorithms will be essential in understanding the working principles of multi-objective evolutionary algo-

rithms. It is also worth mentioning that there exist other evolutionary and nature-inspired algorithms, such as ant-colony optimization, simulated evolution, DNA computing, and cultural algorithm, descriptions of which can all be found in the literature and reviews [50, 51, 52, 53] and textbooks and review of evolutionary algorithms [54, 16, 55, 56, 57, 58] As illustrated in Figure 1.5 the evolutionary algorithms belong to the class of Monte Carlo algorithms and hence are perfectly suited to be converted to parallel computer code. This inherent property of the evolutionary algorithms allows to take full advantage of advanced computer hardware architecture and allows to expand into areas of optimization which have been concealed to date.

1.3.5.1 Evolution Strategy (ES)

As the parent family of evolutionary algorithms, the class of evolution strategies [59, 60, 24, 61], sometimes also referred to as evolutionary strategies, are search paradigms inspired by the principles of biological evolution. The basic concept is a repeated process of stochastic variations followed by selection: in each generation (or iteration), new offspring (or candidate solutions) are generated from their parents, their fitness is evaluated, and the better offspring are selected to become the parents for the next generation as illustrated bellow (Figure 1.15):

The problems typically addressed with evolution strategies are *continuous black-box*

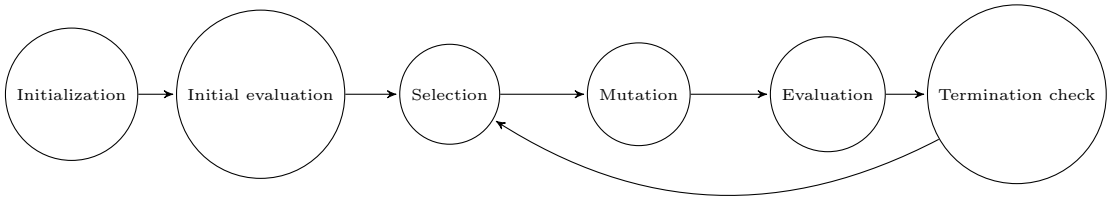


Figure 1.15 – Evolution Strategy illustrated as basic concept

optimization where the search space is the continuous domain $\mathbb{R}^n$, and solutions in search space are $n-$dimensional vectors, denoted as x. The objective or fitness function has the form $f : \mathbb{R}^n \to \mathbb{R}, x \to f(x)$ to be minimized as seen in the previous chapters as a normal single-objective optimization problem. Now, for the evolution strategies, we assume a *population*, P, of so-called *individuals*, where each individual consists of a solution vector $x \in \mathbb{R}^n$ (visible traits or phenotype) and further endogenous parameters, s (hidden traits) and an additional associated fitness value, $f(x)$. The individuals can also be denoted as *parents* or *offspring*, depending on the context. The general procedure for the Evolution Strategies is as follows,

1. Parents are picked from the population (mating selection) and the new off-spring is generated by recombination of the parent generation;
2. The new off-spring undergo random mutation and become new members of the

population;

3. Environmental selection reduces the population to its original size.

Hence, evolution strategies employ the following main principles that are specified and applied in the operators and algorithms it is also attempt to stress the differences to other evolutionary algorithms where possible.

Environmental Selection is used as alleged *truncation selection*. Only the μ best individuals from the population survive based on their fitness $f(x)$. This is in contrast to roulette wheel selection in *genetic algorithms* [54, 16]. In evolutionary strategy only fitness *ranks* are used. Thus, the environmental selection is deterministic in contrast to evolutionary programming and like many other evolutionary algorithms where the environmental selection has a stochastic component.

Mating Selection and Recombination picks individuals from the population to become the new parents. Usually one can distinguish between two common scenarios for mating selection and recombination.

Fitness-Independent mating selection and recombination are either deterministic or stochastic but are not dependent on the fitness values of the individuals. Hence the more dominant process to drive the evolution toward better solutions is the *Environmental selection*.

Fitness-Based mating selection and recombination rely on the rank of the fitness value of the parents in a deterministic way. Therefore, the *Environmental selection* can be neglected in that case.

Mutation and Parameter Control. A random and unbiased mutation process introduces small changes to an individual and usually affect the variables altogether. The impact and consequently the size of these changes depends on the endogenous parameters that are altered over time. These parameters are also referred to as control parameters, or *endogenous strategy parameters*, of which the *step-size* σ is an example. On the other hand, *exogenous strategy parameters* such as the parent number μ are fixed throughout the optimization. The parameter control is the most delicate part of evolution strategies but indispensable and an central feature thereof.

Unbiasedness is the underlying process to introduce new "information" to the variation in the mutation or recombination mechanisms. Hence, it is important that this is the antipode to the selection where a dominant bias towards better solutions occur. Unbiasedness and maximum exploration are in accord and evolution strategies are

unbiased in the following respects.

- The kind of mutation distribution (normal distribution), is preferred since of its rotational symmetry and maximum entropy which corresponds to a maximum exploration [62]. Decreasing the entropy would introduce prior information and hence a bias.
- Endogenous strategy parameters and object parameters are unbiased under mutation and recombination and mutation. Also, mutation usually has expectation zero.
- It is important that invariance properties avoid a bias towards some specific representation of the fitness function or search space. Thus, parameter control in evolution strategies strive for invariance properties [63].

Algorithm 3: evolution strategy pseudocode

given;
$n, \lambda, \mu \in \mathbb{N}_+$
initialization;
$x \in \mathbb{R}^n, P = \{ x\}, s$
while *not happy* **do**
  for $k \in \{1, \ldots, \lambda\}$ **do**
    $s_k = mutate_s(s)$
    $x_k = mutate_x(s_k, x_k)$
  $P \leftarrow P \cup \{(x_k, s_k, f(x_k)\}$
  $P \leftarrow select - \mu - best$

A general pseudocode for the evolution strategy is shown in Algorithm 3. Given is a population P with μ individuals. The individuals are represented as 3-tuple $(x_k, s_k, f(x_k))$ where s_k are the endogenous strategy parameters (control parameters); typically these determine the mutation through the step-size $\sigma \in s_k$ amongst other parameters, the $f(x_k)$ represents the fitness or objective function $f : \mathbb{R}^n \rightarrow \mathbb{R}$ to be minimized and $x_k \in \mathbb{R}^n$ the solution or object parameter as an element of the search space. In every generation, first λ offspring are generated, thus from recombination of the $\rho \leq \mu$ parents from P, followed by mutation of s and x. Accordingly, the new offspring is added to P and finally, the best μ individuals are selected to retain P.

1.3.5.2 Covariance Matrix Adaptation Evolution Strategy (CMA-ES)

The *covariance matrix adaptation evolution strategy* (CMA-ES) [14, 64, 65, 66] can be seen as a de-facto standard in continuous domain evolutionary computation. As we will see, the covariance matrix adaptation evolution strategy is a special case of the more

general algorithm 3. First, we introduce and repeat some nomenclature before the focus will be laid on a very specific case of covariance matrix adaptation algorithm. For a more complete review of the covariance matrix adaptation algorithm we refer the reader to Hansen et al [67] and [68].
Figure 1.16 illustrates an actual optimization run with covariance matrix adaptation on a simple two-dimensional problem [69]. The spherical optimization landscape is depicted with solid lines of equal f-values and the population (dots) is much larger than necessary, but clearly shows how the distribution of the population (dotted line) changes during the optimization. On this simple problem, the population concentrates over the global optimum within a few generations [69].

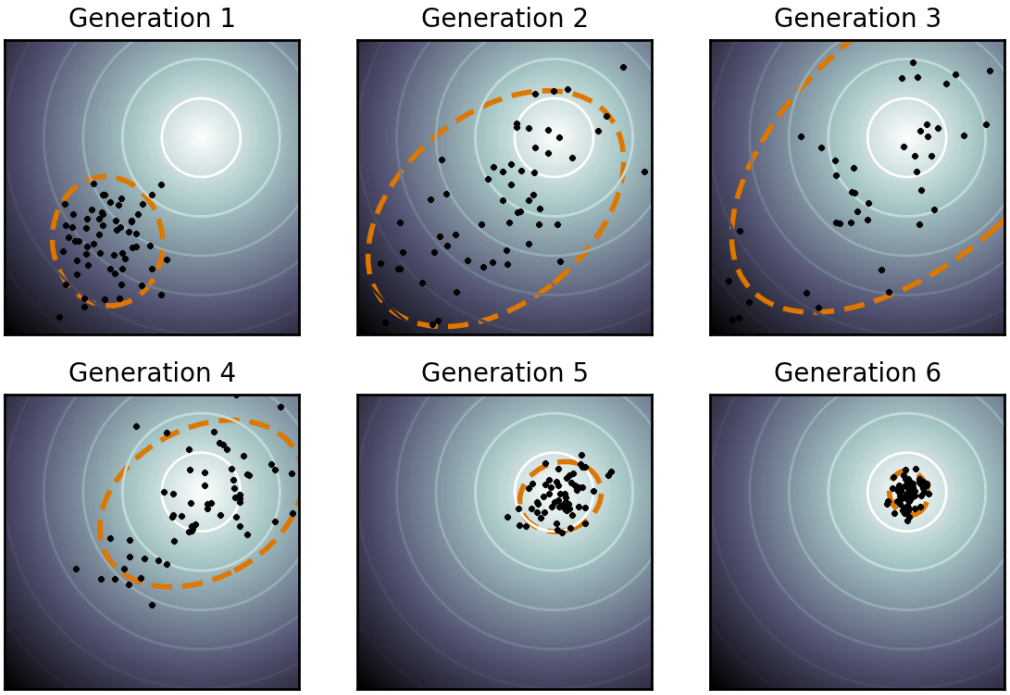


Figure 1.16 – Covariance Matrix Adaptation run for a simple two-dimensional problem [69]

As elaborately described in [67], the well known (μ, λ)-selection scheme for the covariance matrix adaptation algorithm, where the best μ of λ offspring selected from the next parent population and all former parents are discarded, is permanently replaced by the elitist $(1+\lambda)$-selection, that is, the parent population consists of a single individual generating λ offspring and the best individual out of parent and offspring becomes the parent of the next generation. In case where the parent is only generating one $\lambda = 1$ offspring, the selection is termed $(1+1)$. This elitist mechanism is used as selection schema since it showed to be a simple and effective technique [70] indeed.
Figure 1.17 schematically explains the $(1+1)$ selection whereas from x_0 a new solutions

x_1 is created. But since from the total set $\{x_0, x_1\}$ the solution x_0 maintains the better objective function value $f(x_0)$ it is selected to form the new generation. Through the mutation of the endogenous strategy parameters s_k the covariance and/or step-size are modified (dotted red line) and a new solution x_2 is produced. Since x_2 is a better solution than x_0 it is picked as parent for the next generation and again the control parameters are adapted (solid black line).

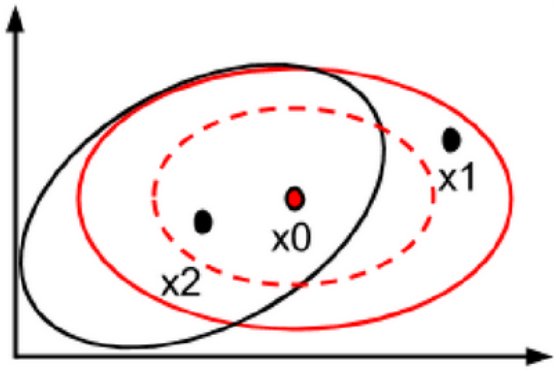


Figure 1.17 – $(1+1)$-Covariance Matrix Adaptation run

Nomeclature In the CMA-ES, each individual, a can be represented as a 5-tuple $a = [x, \overline{p}_{succ}, \sigma, p_c, C]$ with its candidate solution or parameter vector $x \in \mathbb{R}^n$, an averaged success rate $\overline{p}_{succ} \in [0,1]$, the global step size $\sigma \in \mathbb{R}_+$, an evolution path $p_c \in \mathbb{R}^n$, and the covariance matrix $C \in \mathbb{R}^{n \times n}$. The parameters $\{\overline{p}_{succ}, \sigma, p_c, C\}$ can be viewed as the s_k endogenous strategy parameters. In addition, following nomenclature is used [67]:

$f : \mathbb{R}^n \to \mathbb{R}, x \to f(x)$ is the scalar objective function to be minimized. For the MO-CMA-ES, $f : \mathbb{R}^n \to \mathbb{R}^m, x \to f(x)$ is the vector-valued objective function.

$\mathcal{N}(m, C)$ is a multi-variate normal distribution with mean vector m and covariance matrix C. If a random variable x is distributed according to the distribution $\mathcal{N}(m, C)$ it is notated as $x \sim \mathcal{N}(m, C)$

Elitist $(1+1)$-CMA-ES algorithm is described within three routines. In the main routine, the sampling procedure and the update routine in case the solution a_{parent} is updated depending on whether the new solutions is better than a_{parent} according to the $(1+1)$-selection.

Algorithm 4: (1+1)-CMA-ES algorithm

g=0, initialize $a_{parent}^{(g)}$

repeat

  $a_{parent}^{(g+1)} \leftarrow a_{parent}^{(g)}$

  $x_{new}^{(g+1)} \sim \mathcal{N}(x_{parent}^{(g)}, \sigma^{(g)^2} C^{(g)})$

  updateStepSize $\left(a_{parent}^{(g+1)}, \lambda_{succ}^{(g+1)}\right)$

  if $f(x_{new}^{(g+1)}) \leq f(x_{parent}^{(g)})$ **then**

    $x_{parent}^{(g+1)} \leftarrow x_{new}^{(g+1)}$

    updateCovariance $\left(a_{parent}^{(g+1)}, \frac{x_{parent}^{(g+1)} - x_{parent}^{(g)}}{\sigma_{parent}^{(g)}}\right)$

  $g \leftarrow g + 1$

until *stopping criterion is met*;

The step size is updated based on the success rate $p_{succ} = \lambda_{succ}^{(g+1)}$, after sampling the new candidate solutions. That is, $\lambda_{succ}^{(g+1)}$ is either 0 or 1 depending on whether the candidate solution a_{parent} produced a better offspring or not. It implements the well-known

Procedure updateStepSize$(a = [x, \overline{p}_{succ}, \sigma, p_c, C], p_{succ})$

$\overline{p}_{succ} \leftarrow (1 - c_p)\overline{p}_{succ} + c_p p_{succ}$

$\sigma \leftarrow \sigma \cdot exp\left(\frac{1}{d} \frac{\overline{p}_{succ} - p_{succ}^{target}}{1 - p_{succ}^{target}}\right)$

heuristic rule that the step size should be increased if the success rate is high, and the step size should be decreased if the success rate is low [71], whereas c_p is the success rate averaging parameter. The damping parameter d controls the rate of the step size adoption and is chosen $d = 1 + \frac{n}{2}$ [70], where n is the dimensionality of the search space. If the new candidate solution x_{new} was better than the parent individual, the covariance matrix is updated as in the original CMA-ES [14]. The constants c_c and c_{cov} are learning rates for the covariance matrix and evolution path, respectively. The update of the evolution path p_c is coupled to the value of $\overline{p}_{succ}$ and the factor $\sqrt{c_c(2 - c_c)}$ normalizes the variance of p_c [14] viewed as a random variable. Consequently, the evolution path p_c is used to update the covariance matrix as outer product. The new covariance matrix then is a weighted mean of the old covariance and the matrix of the outer product.

Strategy Parameter are target success probability p_{succ}^{target}, step size damping d success rate averaging parameter c_p, cumulation time horizon parameter c_c, and covariance matrix learning rate c_{cov}. For derivation, comparison and detailed explanation see [14, 65, 64, 67, 70, 72]. Default values are given in Table 1.1.

Procedure updateCovariance($a = [x, \overline{p}_{succ}, \sigma, p_c, C], x_{step} \in \mathbb{R}^n$)

if $\overline{p}_{succ < p_{thresh}}$ **then**
  $p_c \leftarrow (1 - c_c)p_c + \sqrt{c_c(2 - c_c)}x_{step}$
  $C \leftarrow (1 - c_{cov})C + c_{cov} \cdot p_c p_c^T$
else
  $p_c \leftarrow (1 - c_c)p_c$
  $C \leftarrow (1 - c_{cov})C + c_{cov} \cdot (p_c p_c^T + c_c(2 - c_c)C)$

Table 1.1 – Default parameters for the (1+1)-CMA-ES

Step size control		
$d = 1 + \frac{n}{2}$,	$p_{succ}^{target} = \frac{1}{5+\sqrt{\frac{1}{2}}}$,	$c_p = \frac{p_{succ}^{target}}{2+p_{succ}^{target}}$
Covariance matrix adaptation:		
$c_c = \frac{2}{n+2}$	$c_{cov} = \frac{2}{n^2+6}$	$p_{thresh} = 0.44$

Initialization of the initial individual, $a_{parent}^{(0)}$ are set to $\overline{p}_{succ} = p_{succ}^{target}$, $p_c = 0$, and $C = I$, where $p_{succ}^{target} = \frac{1}{5+\sqrt{\frac{1}{2}}}$. The covariance C is scaled and normalized given the lower and upper constraints on the variables. The initial candidate solution $x \in \mathbb{R}^n$ and the initial $\sigma \in \mathbb{R}_+$ must be chosen problem-dependent and x can be set uniformly random, given the box-constrains, as well as the initial step-size which can be chosen as $\sigma = \sqrt{Var(x)}$, based on the variance of x.
Another approach usually taken [70, 73] is to perform a hypercube scaling which scales the variables x to the range of $[0 \ldots 1]$. This transformation has also the benefit that for the covariance matrix adaptation very basic initial parameters can be used [70, 70, 73]. Hence the entire optimization process is carried out in the transformed search space which is scaled to the hypercube $[0 \ldots 1]^n$ and back-transformed for the objective evaluation.

1.3.6 Multi-objective Covariance Matrix Adaptation Evolution Strategy

As we have seen, the most prominent feature of the evolutionary algorithms is that a population of solutions is processed in every iteration, even though the population can have size one. This feature alone attributes for a tremendous advantage for its use in solving multi-objective optimization problems. Call to mind Section 1.2 that one of the goals of an ideal multi-objective optimization procedure is to find the true Pareto-front and hence as many optimal solutions as possible. The main idea now is to modify the evolutionary algorithm in a way to so that a population of Pareto-optimal solutions can be found.
Usually one distinguishes between multi-objective optimization algorithms that do not use any elite-preserving operation and the so called Elitist multi-objective optimization

algorithms [16]. Detailed description of the former category can be found in [16] and [74]. As the name suggests, the elite-preserving operator favours the elites of a population by conserving their properties somehow to the next generation. We already encountered a simple example of this mechanism by introducing the $(1+1)$ selection scheme where the offspring is directly compared to the parent and a direct selection is carried out on those two solutions. No matter how elitism is introduced, it assures a certain deterioration and thus it is evident to include it in the evolutionary algorithm [75].
Therefore, based on the (1+1)-CMA-ES we describe the multi-objective evolution strategy according to [75], for a detailed description and a performance evaluation on benchmark functions we refer to [76, 70, 77]. After a brief summary of the considered selection mechanisms, which are based on the non-dominated sorting schema explained in Section 1.2.3 and crowding distance, we finally describe the (1+1)-MO-CMA-ES briefly.

Pareto Optimality. As we have learned in Definition 3, the non-dominated set of the entire search space is termed the Pareto-optimal set. That is, all solutions which are member of the Pareto-optimal front are not dominated by any solutions in the search space.

Multi-objective Selection. First, all elements in a population A of candidate solutions are ranked based on their level of non-dominance. The solutions which are non-dominated in A are denoted by $ndom(a) = \{a \in A \mid \nexists\ a' \in A : a' \prec a\}$ and get rank 1. Recursively, the other ranks are defined by considering the set without the solutions with lower ranks. That is, let $dom_0(A) = A, dom_l(A) = dom_{l-1}(A) \setminus ndom_l(A)$, and $ndom_l(A) = ndom(dom_{l-1}(A))$ for $l \in \{1,...\}$. Formally $a \in A$ is defined as the level of non-dominance $r(a, A)$ to be i iff $a \in ndom_i(A)$.
Since solutions can have the same level of non-dominance, a second sorting criterion is needed for ranking. We considered the crowding-distance [16] mechanism as it is used in the NSGA-II algorithm and schematically presented in Figure 1.18. The basic idea is to rank solutions with the same level of non-dominance A' according to how much they contribute to the spread (or diversity) of objective function values in A'. The crowding-distance for M objectives of $a \in A'$ is given by

$$c(a, A') = \sum_{m=1}^{M} \frac{c_m(a, A')}{(f_m^{max} - f_m^{min})} \tag{1.4}$$

where f_m^{max} and f_m^{min} are estimations of the minimum and maximum value of the mth objective and

$$c_m(a, A') := \begin{cases} \infty, \text{ if } f_m(a) = min\{f_m(a') | a' \in A'\} \text{ or } f_m(a) = max\{f_m(a') | a' \in A'\} \\ min\{f_m(a'') - f_m(a') | a', a'' \in A' \setminus \{a\} : f_m(a') \leq f_m(a) \leq f_m(a'')\}, \text{ otherwise.} \end{cases} \tag{1.5}$$

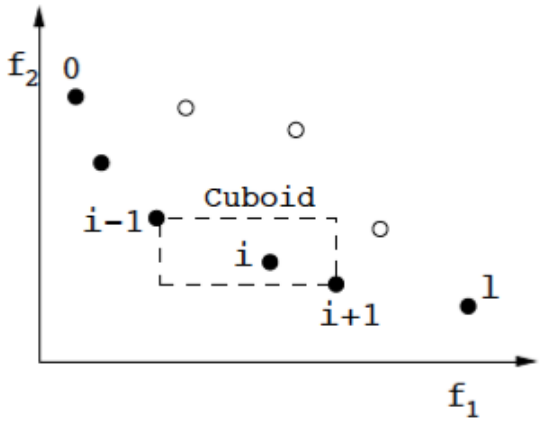


Figure 1.18 – Crowding distance calculation [2]

Given the level of non-dominance and the crowding-distance, the relation is defined as

$$a \prec_{c,A'} a' \Leftrightarrow r(a, A') < r(a', A') \tag{1.6}$$

for $a, a' \in A'$. Hence, a is better than a' when compared using $\prec_{c,A'}$ if either a has a lower (better) level of non-dominance or a and a' are on the same level but a is in a "lesser crowding region of the objective space" and therefore induces more diversity.

Definition 4. a is better than a' when;

- solutions a has a better rank
- If they have the same rank but solution a has a better crowding distance than solution a'

The time complexity of the crowding-distance of N non-dominated solutions is $\mathcal{O}(MNlogN)$ [16]. The crowding distance is related to the spread of solutions, and the basic goal is to get an even distribution of the Pareto front.
Other approaches such as the hyper-volume contribution [78, 79, 77] could also be considered for a second comparison criterion. Additionally, we propose the kernel density estimation, as illustrated in Figure 1.19, for an easy to apply criteria. With this approach one would get an good estimate for the spread of a solution and favour the less crowded area of a front indeed.

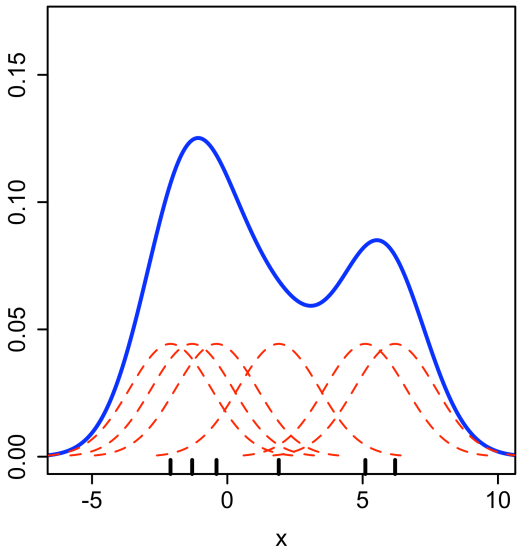


Figure 1.19 – The 6 individual kernels are the red dashed curves, the kernel density estimate the blue curves. The data points are the rug plot on the horizontal axis [23]

MO-CMA-ES. In the λ_{MO}×(1+1)-MO-CMA-ES, a population of λ_{MO} (1+1)-CMA-ES are maintained. The kth individual in the generation g is denoted as the tuple $a_k^{(g)} = [x_k^{(g)}, \overline{p}_{succ}^{(g)}, \sigma_k^{(g)}, p_{c,k}^{(g)}, C_k^{(g)}]$.
The λ_{MO} parents generate in each case one offspring in every generation g. The set $Q^{(g)}$ is composed of the offspring and the parents. The step sizes of a parent and its offspring are adopted depending on whether the mutations were successful, that is, whether the offspring is better than the parent given the relation $\prec_{Q^{(g)}}$. The covariance matrix update of the offspring is then given through the change in parameter space. Both the covariance matrix update and the step size are identical to the single-objective CMA-ES. The best λ_{MO} individuals in $Q^{(g)}$ sorted by $\prec_{Q^{(g)}}$ build the next parent generation.
Hence, the multi-objective optimization algorithm reads:

Algorithm 5: $\lambda_{MO}\times(1+1)$-MO-CMA-ES

g=0, initialize $a_k^{(g)}$ for $k = 1,...\lambda_{MO}$
repeat
  for $k = 1,...,\lambda_{MO}$ **do**
    $a'^{(g+1)}_k \leftarrow a_k^{(g)}$
    $x'^{(g+1)}_k \sim \mathcal{N}(x_k^{(g)}, \sigma_k^{(g)^2} C_k^{(g)})$
  $Q^{(g)} = \{a'^{(g+1)}_k, a_k^{(g)} | 1 \leq k \leq \lambda_{MO}\}$
  for $k = 1,...,\lambda_{MO}$ **do**
    updateStepSize $\left(a_k^{(g)}, \lambda^{(g+1)}_{succ,Q^{(g)},k}\right)$
    updateStepSize $\left(a'^{(g+1)}_k, \lambda^{(g+1)}_{succ,Q^{(g)},k}\right)$
    updateCovariance $\left(a'^{(g+1)}_k, \frac{x'^{(g+1)}_k - x_k^{(g)}}{\sigma_k^{(g)}}\right)$
    for $i = 1,...,\lambda_{MO}$ **do**
      $a_i^{(g+1)} \leftarrow Q^{(g)}_{<:i}$
  $g \leftarrow g + 1$
until *stopping criterion is met*;

Here $\lambda^{(g+1)}_{succ,Q^{(g)},k}$ is defined as 1

$$\lambda^{(g+1)}_{succ,Q^{(g)},k} = \begin{cases} 1, \text{ if } a'^{(g+1)}_k \prec_{Q^{(g)}} a_k^{(g)} \\ 0, \text{ otherwise} \end{cases} \tag{1.7}$$

if the parent $a_k^{(g)}$ has successfully produced a better offspring, otherwise it is set to 0. Additionally, $Q^{(g)}_{<:i}$ is the ith best offspring in $Q^{(g)}$ w.r.t $\prec_{Q^{(g)}}$
The Box constraints are handled such that the solution x is taking the upper or lower limit if the boundaries are violated. More sophisticated approaches such as penalty functions [80] or covariance manipulation [65] also exist but have currently not been implemented.

2 Semiempirical Methods

In this chapter, the theoretical quantum chemical framework employed is presented. The aim is to give a general introduction to the semi-empirical methods and the essentials of modern semi-empirical molecular orbital theory. However, the focus is set on the Neglect of Diatomic Differential Overlap (NDDO) and Density Functional Tight Binding (DFTB) methods and therefore their theoretical basis is elaborated. More details of the topics can be found in standard textbooks [6, 81, 7, 8, 9, 82] or in the review papers cited in different sections of this chapter [83, 84, 85, 86, 87].

2.1 Introduction

Quantum mechanics provides the conceptual framework for comprehending chemistry and the theoretical foundation for computational methods that model the behaviour and electronic structure of chemical compounds. In contrast to the classical force field approach, where a functional form with parameter sets are applied to model compounds, this crude approximation of molecular modelling does not allow to describe chemical bond breaking and are challenged to account for polarization of the environment besides other deficiencies [88, 89, 90, 91]. Despite its drawbacks, the molecular mechanics methods are very successful in describing systems sizes which exceed the capabilities of quantum chemistry approaches or simulation time necessary to observe certain event occurrences. Figure 2.1 illustrates simulation time and system sizes available with given methods.

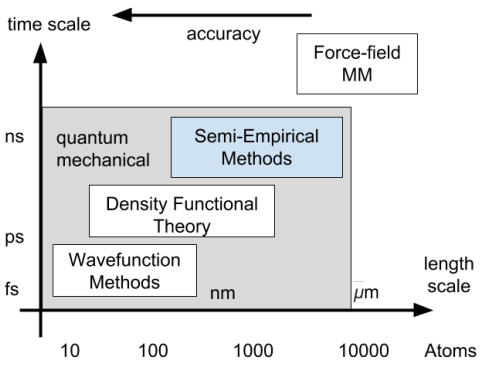


Figure 2.1 – Different time and length scales accessible with different computational methods.

Nevertheless, since for a vast majority of chemical problems and thus properties, the electronic structure plays a crucial role, one will not be able to avoid a quantum mechanical approximation. In principle one can distinguish three types of approaches in quantum chemistry:

- Wave-function methods
- Density functional theory
- Quantum-chemical semi-empirical methods

The Wave-function approach provides a convergent path to the exact solution of the Schroedinger equation. However, because of its complexity those methods are very costly and therefore restricted to relatively small molecules [92]. Density functional theory (DFT) has developed to one of the most popular computational chemistry approaches because of its favourable price/performance ratio, enabling accurate calculations on fairly large systems [93, 92, 94]. The biggest disadvantage of the density functional theory methods is that there is no systematic path of improvement despite its first-principle characteristics. The simplest variant of electronic structure theory are termed Quantum-chemical semi-empirical methods since they involve integral approximations and parametrizations that limit their accuracy but enable the access of larger systems and longer molecular dynamics time scales and thus bridging the gap between molecular mechanics and ab initio calculations. All these semi-empirical methods employ a simple strategy. The first-principle formalism is the base on which rather drastic assumptions and simplifications are proposed. Typically one neglects many of the less important terms in the underlying equations allowing a drastic speed up of the resulting calculations. The errors introduced through those approximations are expected to be compensated through empirical parameters which are incorporated into the formalism and fitted

against reference data. Needless to say that it depends on the approximations and thus semi-empirical method, if a chosen model retains the essential physical properties of interest and it can well be that very important properties are getting lost on the way. Hence it is important for a systematic benchmarking and validation to explore the limits of any given method and underlying model. In this chapter, we consider semi-empirical methods that are based on the molecular orbital (MO) theory, which was proposed in the early twentieth century and revolutionized the study of bonding by approximating the positions of bonded electrons (molecular orbitals) as Linear Combination of Atomic Orbitals (LCOA). First a brief historical overview and basic concepts are given followed by a short summary of two major methods in semi-empirical approaches

2.2 Historical Overview

Hückel proposed in 1930 one of the earliest semi-empirical approaches in quantum chemistry which was called pi-electron or Huckel method. It is a very simple linear combination of atomic orbitals for the determination of energies of pi electrons in conjugated hydrocarbon systems, such as ethene, benzene and butadiene, see section 2.3.2. Later, it was extended to conjugated molecules such as pyridine, pyrrole and furan which contain other atoms than carbon. The extended Hückel method [95] developed by Roald Hoffmann [96] since 1963 is based on the Hückel method but also includes the sigma orbitals, thus describing all valence electrons. It has been applied in many qualitative studies of organometallic and inorganic compounds. These inceptions of semi-empirical methods have had a lasting impact on physical chemistry since they guided the thinking and development of qualitative molecular orbital theory, and hence employed a way for rationalizing chemical phenomena in terms of orbital interactions.
Those early type methods only included one-electron integrals and therefore are non-iterative and mathematically simpler to solve. The semi-empirical self-consistent field (SCF) methods on the other hand are taking into account the two-electron integrals explicitly. The very first of such approaches was also restricted to pi electrons and based on the PPP (Pariser-Parr-Pople) formulation [97]. The resulting method is based on the zero differential overlap (ZDO) approximation which found its way into many succeeding semi-empirical methods and could reliably describe the electronic spectra of unsaturated molecules [98]. The generalisation to valence electrons was first proposed by Pople [99] by explicitly including the electron-electron repulsion terms, but neglecting many of them and approximating some. The modelling of the integral terms, that is the neglect of the differential overlap, obey rotational invariance and other consistency criteria and give rise to different approximations which are hierarchical specified; complete neglect of differential overlap (CNDO), intermediate neglect of differential overlap (INDO) [100] and neglect of diatomic differential overlap (NDDO) methods (see section 2.3.3.1).
The parametrization of those earlier methods was carried out to reproduce ab initio Hartree-Fock (HF) results with a minimal basis set [5]. Hence, those earlier calculations

could at best reach the accuracy of the target HF methods, which meanwhile are known to be rather poor. Nevertheless, the original CNDO/2 method [101] and the development thereof, the INDO approach, are the most prominent example of early parametrizations introduced by Pople [99, 102]. The methods are now rarely used in their original form with some exceptions but are basis for several other methods. Indeed, Dewar et al. pursued another parametrization strategy aiming at a realistic description of ground-state potential surfaces and used experimental reference data. This path culminated in the MINDO/3 [103] method which was based on the INDO approach and the MNDO [104, 105] and AM1 [106] methods which grounded on NDDO approximations. In turn, a different parametrization of the MNDO model resulted in the PM3 [107] method, which is formally very close to AM1 but differs in the core repulsion function and the RM1 variant which is a re-parametrization of AM1 with a much larger set of reference data yielding improved results.

There are two other noteworthy INDO-based approaches in the time before 1990. The SINDO1 method [108] developed by K. Jug and co-workers [109] includes symmetric orthogonalised one-electron integrals and d orbitals for atoms of the second row of the periodic table. The SINDO parametrization was carried out against ground-state properties and the method performs better for hypervalent compounds than other semi-empirical methods. A further development and modification lead to the MSINDO method [110]. A different direction was followed by Zerner and co-workers in the development of the INDO/S method [111, 112] that targets the calculation of electronic spectra, in particular vertical excitation energies, using configuration interaction with single excitations (CIS). INDO/S was parametrized at the CIS level and turned out to be rather successful in spectroscopic and related areas.

In the time since 1990, the MNDO model was extended to include a spd basis [113, 114], which in turn enabled the treatment of heavier elements and led to improved MNDO/d results. The extension to a larger spd basis was embraced and also used for the latest general purpose parametrization such was PM6 and PM7 [33, 115]. Those two NDDO based methods cover essential the entire periodic table and hence are successfully applied to compute molecular and solid-state properties. Another general-purpose parametrization of the MNDO model employing a functional group-specific modification of the core repulsion function led to the PDDG/MNDO and PDDG/PM3 variants [116]. The Pairwise Distance Directed Gaussian (PDDG) modification apparently provides good description of the van der Waals attraction between atoms, and the PDDG/PM3 model appears to be well suitable for calculations of intermolecular complexes. Hence, the modification of the empirical core repulsion function in the MNDO model was adopted in various special-purpose parametrizations.

Exceeding the MNDO model, some orthogonalization models (OM1, OM2 and OM3) have been proposed that contain orthogonalization correction in the one-electron terms of the NDDO Fock matrix to account for the effects of Pauli exchange repulsion [117, 118]. The explicit representation of Pauli exchange repulsion in the OMx families was shown to improve the description of the electronically excited states, conformational properties,

and non-covalent interactions.
Conceptually, the semi-empirical methods mentioned above are models to simplify ab initio molecular orbital approach. On the other hand, a semi-empirical tight-binding (TB) approach of DFT was also developed and termed DFTB [119]. The method was extended to a self-consistent charge model (SCC-DFTB) [120]. Despite the conceptual origin and the derivation of the DFTB approaches seem different from those of conventional semi-empirical procedures, they share many similarities not only in implementation and actual computational procedures. The DFTB method embrace several integral approximations and extensive parametrization in a way such that it is appropriate to consider them as semi-empirical methods on par with the traditional ones [84].
In the present chapter, we focus on the methods that are nowadays of relevance since the role of semi-empirical calculations has changed during the past 40 years. Historically, the MNDO-type methods (namely MNDO,AM1 and PM3) served as workhorse for quantum-chemical computations and at present, these methods are still widely used, along with more recent versions like PM6. As DFT gained traction in the ab-inito community, the semi-empirical approach thereof DFTB entered the field of biochemistry and material science and was applied successfully and remains popular. Therefore we concentrate on the MNDO-type and DFTB approaches in the following.

2.3 Methods

In this section, we provide an review on the beginnings of the semi-empirical methods as well as MNDO and DFTB approaches. For a broader description of the formalisms, the reader is advised to consult several comprehensive review articles and publications.

2.3.1 Basic Concepts and Theory

A semi-empirical *model* or *scheme* is defined by the underlying theory and the modalities of approximation and hence the resulting interaction that is included by the model or the level of integral approximations. On the other hand, the resulting *method* is based on a model but depends on the implementation and parametrization implied on the model. It should be mentioned that although methods and models are related, there may be large differences and caution is advised by drawing conclusion from a method to a model. For example, the AM1 and PM3 method are both based on the MNDO model but give different results for different properties. Hence, to explore the boundaries of a model in principle all limits of the entire methods and parametrizations thereof would need to be explored.

All the MNDO models employ a Hartree-Fock SCF-MO treatment with a minimal basis set for the valence electrons and the core electrons are accounted through a reduced nuclear charge. Dynamic correlation effects are included in an average approach by a suitable representation of the two-electron integrals and the overall parametrization.

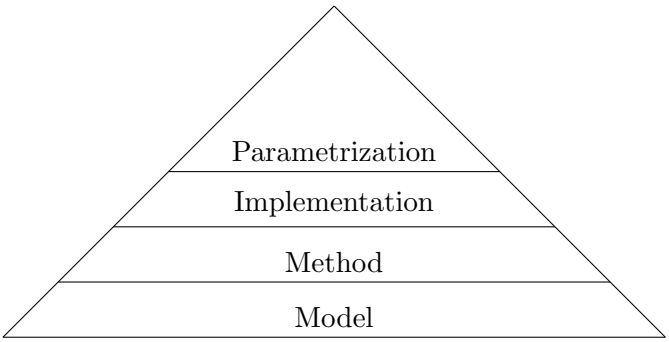


Figure 2.2 – Different layers from model to parametrization

The standard equations are reduced by the integral approximations to neglect all three-center and four-center two-electron integrals. The CNDO model uses zero-differential overlap for the two-electron integrals, whereas INDO includes the integrals that are over orbitals centred on the same atom (one-center two electron integrals). Finally, NDDO adds all two centre integrals for repulsion between a charge distribution on one center and a charge distribution on another centre. In that way, NDDO retains the higher multi-poles of charge distributions in the two-center interactions and hence accounts for anisotropies in these interactions. As such, the approximations are applied to all integrals that involve Coulomb interactions, and to the overlap integrals that appear in the HF secular equations. The integrals are either determined directly from experimental data, calculated exactly from the corresponding analytic formulas or expressed parametrically and fitted thereafter. Resolving from experimental data usually is only possible for the one-center integrals, where atomic spectroscopy data is available. The introduction of a parametric expressions need careful analysis of the corresponding analytic integral and is a difficult and challenging procedure. The decision on how the integrals are represented is a mayor part of the *implementation* of a model and therefore an important component of the resulting method. The remaining task is to *parametrize* the implementation to determine optimal values from a given set of reference data. Commonly, experimental or high-level ab initio data are used as reference set and different optimization algorithms with diverse functions were applied. It has to be noted that the quality of the results strongly depend on the parametrization and major difficulties remain to guarantee parameter transferability for different setups. One should be aware that it is challenging to estimate the limits of a parametrization since the reference set only covers a certain scope. For example, could the parametrization of Hydrogen vary significantly for different chemical environments. Hence, caution is advised when assessing methods based on parametrizations since it is not clear whether the limitations are originated on the level of the model, implementation or parametrization.

2.3.1.1 LCAO-SCF

As mentioned, the essence of modern semi-empirical molecular orbital theory is based on the two approximations: self-consistent field (SCF) [121] and linear combination of atomic orbitals (LCAO) [122]. There are elaborate descriptions available in many standard textbooks [87, 7, 6, 8] but for understanding MNDO-like approaches the comprehension is crucial. We can write the Hamiltonian for a molecule that consists of M nuclei and N electrons as

$$H = \sum_{i=1}^{N} \frac{1}{2}\nabla_i^2 + \sum_{A=1}^{M} \frac{1}{2M_A}\nabla_A^2 - \sum_{i=1}^{N}\sum_{A=1}^{M} \frac{Z_A}{R_{A_i}} + \sum_{i=1}^{N}\sum_{j>i}^{N} \frac{1}{r_{ij}} + \sum_{A=1}^{M}\sum_{B>A}^{M} \frac{Z_A Z_B}{R_{AB}} \tag{2.1}$$

where the indices i and j run over the electrons and A and B over the nuclei. The separate terms that make up the Hamiltonian are defined in Table 2.1.

Table 2.1 – Definitions of the Individual Terms in Eq. 2.1

Term	Definition	Variable
$\sum_{i=1}^{N} \frac{1}{2}\nabla_i^2$	Kinetic energy of the electrons	∇_i = the first derivative of the position of electron i with respect to time (velocity)
$\sum_{A=1}^{M} \frac{1}{2M_A}\nabla_A^2$	Kinetic energy of the nuclei (zero within the Born-Oppenheimer approximation)	∇_A = the first derivative of the position of the nucleus A with respect to time (velocity)
$\sum_{i=1}^{N}\sum_{A=1}^{M} \frac{Z_A}{R_{A_i}}$	Nucleus-electron attraction	Z_A is the nuclear charge of atoms A and R_{A_i} is the distance between atom A and electron i
$\sum_{i=1}^{N}\sum_{j>i}^{N} \frac{1}{r_{ij}}$	Electron-electron repulsion	r_{ij} is the distance between electrons i and j
$\sum_{A=1}^{M}\sum_{B>A}^{M} \frac{Z_A Z_B}{R_{AB}}$	Nucleus-nucleus repulsion (constant within the Born-Oppenheimer approximation)	R_{AB} is the distance between atoms A and B

As commonly implied, we make use of the Born-Oppenheimer approximation [123], thus the nuclei moves much slower than the electrons than the former can be regarded as being stationary. This results in the fact that the nucleus-nucleus repulsion can be regarded as a constant and can be neglected in the electronic Hamiltonian:

$$H = H_{nuclear} + H_{electronic} = H_{nuclear} + \sum_{i=1}^{N} \frac{1}{2}\nabla_i^2 - \sum_{i=1}^{N}\sum_{A=1}^{M} \frac{Z_A}{R_{A_i}} + \sum_{i=1}^{N}\sum_{j>i}^{N} \frac{1}{r_{ij}} \tag{2.2}$$

and thus the total Hamiltonian H can be separated into nuclear and electronic component. Hence, the total energy can be written as the sum of the nuclear repulsion energy and

the electronic energy defined by the Hamiltonian $H_{electron}$:

$$E_{total} = E_{electronic} + \sum_{A=1}^{M} \sum_{B>A}^{M} \frac{Z_A Z_B}{R_{AB}} \tag{2.3}$$

The electronic energy is obtained solving the Schroedinger equation [3] from the electronic wave-function. The wave-function $\Phi_{electronic}$ on the other hand is a function of the position and spins of the N electrons of the system:

$$\Phi_{electronic} = \Phi(x_1, x_2, x_3, ..., x_N) \text{ where } x_i = \{r_i, \omega_i\} \tag{2.4}$$

Where r_i represents the position of electron i and ω_i its spin. Evidently, the wave-function is a function of $4N$ variables. Since the Schroedinger equation can only be solved for one electron, we have to introduce approximations. The first of these is the *mean-field* (or *Hartree-Fock*) approximation [4, 124]. Essentially, rather than solving the complete Schroedinger equation for many particles, we approximate the many-particle solution in terms of many one-electron wave-functions, which are solvable. Hence, the approximation writes as:

$$H_{electronic} \approx \sum_{i=1}^{N} h_i \tag{2.5}$$

where h_i is the one-electron Hamiltonian for electron i. This results in the Hartree product Φ^{HP}, which is an approximation for a many-electron wave-function $\Phi_{electronic}$

$$\Phi^{HP}(x_1, x_2, ..., x_N) = \chi_1(x_1)\chi_2(x_2)\cdots\chi_N(x_N) \tag{2.6}$$

in Eq. (2.6) χ_i are the spin orbitals, which are one-electron wave-functions. The Schroedinger equation based on the Hartree approximation can be written as

$$H\Phi^{HP} = E\Phi^{HP}$$

so that the eigenvalues ϵ_i of the one-electron wave-functions χ_i can be summed to give the electronic energy:

$$E_{electronic} = \sum_{i=1}^{N} \epsilon_i$$

Because electrons are fermions, they must obey the Pauli exclusion principle [125], which in terms can be formulated as the antisymmetry principle and that states that the wave-function must be antisymmetric with respect to the exchange of any two electrons. Fock pointed out that the Hartree product does indeed not obey the antisymmetry principle. Slater then proposed that the wave-function suggested by Fock can be expressed as a

determinant, Φ^{Slater} [126]

$$\Phi^{Slater} = \frac{1}{\sqrt{N!}} \begin{vmatrix} \chi_1(x_1) & \chi_2(x_1) & \dots & \chi_N(x_1) \\ \chi_1(x_2) & \chi_2(x_2) & \dots & \chi_N(x_2) \\ \vdots & \vdots & \dots & \vdots \\ \chi_1(x_N) & \chi_2(x_N) & \dots & \chi_N(x_N) \end{vmatrix} \tag{2.7}$$

Commonly, this expression is known as the Hartree-Fock (or SCF) wave-function. Nevertheless, one needs to find an expression for the spin orbitals χ_i. This is where the almost universal LCAO approximation, introduced by Erich Hückel [127], comes into play. Hückel s proposal was that molecular orbitals (χ_i) can be represented as a linear combination of atomic orbitals. For a system constituted of N atomic orbitals (AOs)

$$\chi_i = \sum_{j=1}^{N_{AOs}} c_j^i \varphi_j$$

where c_j^i is the coefficient of atomic orbital φ_j in molecular orbital χ_i, and the coefficients are normalized. Because one still cannot solve for the wave-function directly, the variational principle is used to find the correct wave-function with the lowest energy.

2.3.1.2 Implications of LCAO-SCF Theory

Although the LCOA-SCF theory is very successfully applied, it has two serious limitations that needed to be pointed out. The first is a consequence of the SCF approximation and also known as *electron correlation*.
Physically, the simplification of the Hartree product means that the electrons do not feel each other individually rather each electron feels the average electron density of the others. Thus, individual electrons can not avoid each other instantaneously, which they would given their negative charge. Hence, the SCF approximation means that the electron-electron repulsion is overestimated. This effect, as pure consequence of the SCF approximation, is also know as the *dynamic correlation* [128].
The second implication of the LCAO-SCF approximation deals with the limitations placed on the wave-function by the atomic orbitals used to form the MOs. Although the LCAO approximation is very intuitive and provides qualitative understand of bonding effects [129], it nevertheless has no physical basis. Even though it is very convenient for calculations, numerical grids or combination of non-atom-centred functions could be used to describe MOs instead. A serious limitation however is that only wave-functions that are a linear combination of atomic orbitals can be described. The atomic orbitals usually are also refered to as basis set in ab initio and density function theory calculations. Prevailing MNDO-type semi-empirical methods use single-valence basis sets resulting in a representation of the valence shell atomic orbital by only one basis function. Thus, the size of the orbital is fixed, even though in reality some valence orbitals are more or less

diffuse than others.

2.3.1.3 Parametrization

Common to all the semi-empirical methods is that they contain adjustable parameters. Methods based on the same model may differ in the values of these parameters only. Provided that the model to wit the set of approximations is sufficiently flexible and physically realistic, the accuracy of a semi-empirical method depends on only two quantities; the *accuracy* and *range* of the reference data used for parametrization and the *thoroughness* of the optimization procedure. As we shall see in the last chapter, those two properties are not to be underestimated.

Data The set of reference data used in parametrization must satisfy several criteria: Clearly, it must be as accurate as possible and represent the targeted system and properties thereof. It is not obvious how to select the reference data in advance. In the last chapter we will give the example of bulk water as target system and the quest of finding the right reference data to wit answering the question if it is possible to reproduce liquid water properties from cluster reference data.
Several useful collections of reference data are available, such as the NIST database of atomic energy levels [130], reference heats of formation [131], and atomic and molecular ionization potentials [132], and the Cambridge Structural Database [133] for molecular geometries. Despite the vast amount of available experimental reference data, important gaps or deficiencies exits. Where data are missing or are incomplete , one can use reference data generated from high-level theoretical calculations. Of course, great care must be taken relining on the accuracy of all calculated reference data, since the objective is to model the real world.
There is no explicit literature addressing the relation of reference data, parametrization and target system. It is crucial not only for the understanding of the target system but also for the parametrization and for the underlying semi-empirical model that more time and effort is invested to examine this context. Additionally, more systematic approaches are needed as selection process for reference data and given the computational power nowadays more resources should also be spend in investigating the impact and relation thereof.

Parametrization Techniques and Strategies In principle one has the choice between empirical reference sets (i.e geometries from experiments and energies from measurements) or ab-inito calculations. The former approach was advocated for the MNDO parametrization. Thus, the disadvantage is that usually there is fewer data available as for standard computed values as state above.
The main decision on parametrization strategies is the inclusion of objectives. Hence, should geometries, forces, energies, dipoles and much more be included in the optimiza-

tion procedure? Commonly, people applied a scalarization approaches to combine these properties [107, 106, 33]. Thus, the important question was how these objectives are assembled and then optimized. Therefore, arbitrary normalization and weighting were employed and then standard optimization algorithms (such as BFGS [134]) used to obtain final parameter sets. When in the beginnings of the semi-empirical methods mostly geometries and energies played a crucial role, recently the inclusion of forces became popular. These techniques are usually referred to as 'force-matching' algorithm and high accurate ab-inito force and energy calculations are used as reference systems.

2.3.2 Hückel Molcular Orbital Theory

The Hückel Molecular Orbital (HMO) method is a simple yet powerful approach towards qualitative understanding of physical properties such as stabilities and chemical reactivities of organic π-systems. Although the HMO results are admittedly crude and approximate, they are capable of explaining and predicting a large amount of interesting chemistry and are in good agreement with either chemical experience or intuition.
Assuming that the total wave function for poly-electronic systems can be factored into sets of independent, non interacting electronic systems Φ, each of which specify the behaviour of a particular set of electrons. Considering molecules where electrons can be described as either being σ or π-electrons, then it follows that

$$\Psi_{polyelect} = \Phi_\sigma \Phi_\pi \tag{2.8}$$

and further

$$E_{tot} = E_\sigma + E_\pi \tag{2.9}$$

is valid approximation for systems such as hydrocarbon systems given that σ-systems in such cases are independent of the π-system, which usually is referred to as sigma-pi separability. It can be justified by the orthogonality of σ and π orbitals in planar molecules and hence also limited to planar systems. In addition, only π-electron molecular orbitals are included, because these determine the general properties of the particular molecules. A further assumption is that Φ_π can be expressed as a very simple linear combination of atomic orbitals (LCAO see detailed explanation in subsection 2.3.1.1). Hence, the system can be written as

$$\Phi_\pi = \prod_j \psi_j \tag{2.10}$$

where each ψ_j is of the form

$$\psi_j = C_{j1}\phi_1 + C_{j2}\phi_2 + C_{j3}\phi_3 + \cdots + C_{jn}\phi_n = \sum_{i=1}^{n} C_{ji}\phi_i \tag{2.11}$$

Within that scope, the Schrödinger equation and simplifications based on orbital symmetry are introduced and solved accordingly. The Hückel method can also be derived from the Ritz method with a few further assumptions concerning the overlap Matrix S and the Hamiltonian matrix H. Hence, the overlap matrix S is assumed to be the identity matrix meaning that the orbitals are considered orthogonal and there is no overlap between the orbitals. The Hamiltonian matrix $H = H_{ij}$ is parametrised in the following way:

- $H_{ij} = \alpha$ for C atoms and $\alpha = h_A\beta$ for atoms A.
- $H_{ij} = \beta$ if the two atoms are next to each other and both C, and $k_{AB}\beta$ for other neighbouring atoms A and B
- $H_{ij} = 0$ in any other case

As one can see, the Hückel molecular orbital method is a very powerful educational tool, and the method's details do appear in many chemistry textbooks [6, 81, 7].

2.3.3 Neglect of Diatomic Differential Overlap

The NDDO approximation is the foundation for most of the modern (MNDO-like) semi-empirical MO theories. Apparently, the NDDO approximation appears to be extremely robust and does not lead to identifiable systematic errors, as have been identified by previous approximations (INDO). In the Hartree-Fock theory, calculating the electron-electron repulsion requires all integrals of the type $(\mu\nu|\lambda\sigma)$ to wit all integrals in which the indices μ, ν, λ and σ vary from 1 to N_{AO} are needed. This is why the result is very large number of integrals (formally $N_{AO}^4/8$) that need to be calculated and processed every iteration of the SCF procedure. The NDDO approximation reduces all integrals $(\mu\nu|\lambda\sigma)$ to zero in which either atomic orbitals μ and ν or λ and σ are on different atoms. The combinations $\mu\nu$ and $\lambda\sigma$ are known as charge distributions, so that the NDDO approximation can also be expressed as meaning that we only consider integrals between charge distribution $\mu\nu$ and $\lambda\sigma$ situated on single, but no necessarily the same, atoms. Therefore, NDDO reduces the problem of calculating and using the two-electron integrals (electron-electron repulsion) from one of four centres to one of only two. Essentially, we calculate only one- and two-center two-electron integrals ignoring three- and four-center two-electron integrals.
Clearly, one needs an efficient technique to calculate the reaming number of integrals. Usually in ab-initio and DFT calculations the use of basis sets are employed. These basis sets are based on Gaussian functions since these are particularly suitable for calculating the integrals. Gaussian orbitals have the form:

$$\phi_l^m(r) = Y_l^m e^{-\zeta r^2} \tag{2.12}$$

where Y_l^m is the angular part (spherical harmonic function) of the orbital with principal quantum number l and angular momentum quantum number m. The reaming expression $e^{-\zeta r^2}$ specifies the radial behaviour of the wave-function, where ζ is the exponent that describes how fast the wave-function falls off with increasing distance r from the nucleus. Despite their almost universal use as atom-centred basis sets in ab-initio and DFT calculations, Gaussian functions are far from ideal. Since the distance from the nucleus is square in the exponent, the wave-function falls off faster than it should and also describes the wave-function at the nucleus not correctly. A better take would be to use Slater orbitals, with the form

$$\phi_l^m(r) = Y_l^m e^{-\zeta|r|} \tag{2.13}$$

However, given that the two-electron integrals for Slater orbitals are very expensive to calculate, they are seldom used despite their inherent advantages.
Most MNDO-type methods use Slater-type orbitals, but must therefore resort to a fast, approximate technique for obtaining the two-electron integrals. This is the multi-pole approach introduced with MNDO [104]. The idea is that the interaction between Slater orbitals are approximated as interaction between electrostatic monopoles, dipoles, and quadrupoles, which allows the integrals to be calculated very effectively and with the required accuracy.
In standard MNDO-like theories, an important approximation is that the basis set (atomic orbitals) is assumed to be orthogonal to wit the orbitals have zero overlap with each other. The benefit of this assumption is that the initial orthogonalization step in the SCF calculation can be saved, which would slow semi-empirical calculations considerably. Yet, one of the most difficult topics in semi-empirical theories is the treatment of the nucleus-nucleus repulsion. Although in Equation 2.1 and Table 2.1 the Coulomb repulsion seems to be of a simple form, in fact in MNDO-type treatment this is a fairly complex entity. The problem arises from the fact that not all of the Coulomb interactions are treated equally. More explicit, the nucleus-nucleus repulsion is treated exactly according to Equation 2.1 but introducing the NDDO approximation results in some neglect of Coulomb terms involving the electrons. More specifically, the long-range behaviour both of the electron-electron and the nucleus-electron integrals is not correct. Hence, the physically correct nucleus-nucleus interaction in term Equation 2.1 would lead to a net repulsion between neutral atoms or molecules at distance outside their van der Waals radii. Hence, an artificial screening effect needs to be introduced to overcome those shortcomings. The MNDO methods usually employ the nucleus-nucleus repulsion E_{AB} in the following form

$$E_{AB}^{MNDO} = Z_A Z_B (s_A s_A | s_B s_B)(1 + e^{-\alpha_A R_{AB}} + e^{-\alpha_B R_{AB}}) \tag{2.14}$$

where the two constants α_A and α_B are adjustable parameters specific to the elements A and B respectively and the electron-electron integrals are treated accordingly. However, with the nucleus-nucleus repulsion taking the form of Equation 2.14 one is not able

to reproduce hydrogen bonds [135] likely attributed to the repulsion being too strong. Therefore, this term was modified by the addition of up to four Gaussian terms in MNDO/H. Nevertheless, the use of these additional correction is not without hazard because they can lead to spurious minima [136] and introduces a large number of additional parameters which is undesirable. A more practical solution was found by introducing two-center temrs in to the nucleus-nucleus repulsion, as suggested originally for AM1 by Voityuk and Rosch [137]. The term of the nucleus-nucleus repulsion becomes

$$E_{AB}^{MNDO} = E_{AB}^{MNDO}(1 + \delta_{AB} e^{-\alpha_{AB} R_{AB}}) \tag{2.15}$$

where δ_{AB} and α_{AB} are parameters specific to the pair of elements AB.
The fundamental problem, that the Coulomb interactions are not treated equally remains with all suggested corrections to the nucleus-nucleus repulsion. Additionally, the inclusion of a two-center potential can adversely affect the parametrization of other such interactions because the effects of the two potentials are not independent of each other. Whenever possible, two-center potential should be avoided because they are intrinsically problematic. This becomes obvious when parametrizing for the periodic table, since not only atomic parameters but also all atomic pairs need to be parametrized.

2.3.3.1 MNDO

The oldest of the NDDO methods is the in 1977 published MNDO [104, 105]. It was a major improvement over the at that time popular MINDO/3 [103] in accuracy and efficiency. Two main reasons can be accounted for the increase in accuracy: First, a semi-empirical method could represent the Ion-pair/Ion-pair interaction of the type found in hydrazine and hydrogen peroxide. And second, reference data based on experimental results for molecular systems were used in the parametrisation process. Nevertheless, since the method was applied to more and more species, various systematic errors became apparent, the most serious being that hydrogen bonds could not be reproduced.

2.3.3.2 AM1

Covalent bonds are stronger than hydrogen bonds which can be best represented by three terms: an electrostatic, a covalent, and a third term variously called instantaneous correlation, dispersion, or van der Walls interaction. MNDO included the covalent and electrostatic terms, but left out the VDW interaction. To incorporate the effect of the VDW term, during the development of AM1 the core-core interaction in MNDO was modified by the addition of simple Gaussian functions to provide a weak attractive force. The intention was that this extra stabilization would allow hydrogen bond to form. Parameters for H, C, N, and P were optimized using a larger set of reference data, and the resulting AM1 methods was published in 1985 [106]. It has to be noted that over

the following few years, parameters were optimized for more elements but with fixed parameter sets of the previous elements. Needless to say, that this resulted in a piecemeal method - where the parameters depended on the sequence in which the parametrization was carried out. Hence, the parametrization process lacked a systematic approach which unfortunately continues to this day.
Two different philosophical approaches were explored that time carrying out parameter optimization. One, advocated by J. Steward, was to provide the parameter optimization procedure with a wide range of reference data, in the hope that if enough data were provided, the underlying rules of chemistry would be picked up by the parameter optimization. The other approach, advocated by Michael Dewar, used chemical knowledge to guide the progress of the optimization. The size of the training set could be kept minimal through careful selection of the reference data. As Dewar had an encyclopaedic knowledge in this field, his approach had obvious merit and therefore was used in the development of AM1.

2.3.3.3 PM3

In contrast to the parametrization approach taken in AM1, a huge amount of reference data was used in the training set for the development of PM3 [107, 138]. Initially, 12 elements were optimized simultaneously. In contrast to AM1, no external constraints based on chemical experience were applied in the process. Comparing PM3 with AM1, it was found that the average errors for common properties, such as heats of formation, and that it follows that PM3 outperforms AM1. The predictive power of PM3 versus AM1 on the other hand is very difficult to answer. Maybe because of this, PM3 was never as widely used as AM1. PM3 was extended to include most [139], and finally all [140], of the main group elements. Nevertheless, the same problematic remains for the extension of PM3 that the later parametrization were carried out using fixed values for the elements that had previously been parametrized.

2.3.3.4 PM6

In the beginning of 2000, Voityuk and Rosch [137] attempted to improve the accuracy of a method for modelling systems containing molybdenum and hence proposed using diatomic core-core parameters [ref section NDDO]. Initially, this modification was utilized using various pairs of elements in the first PM3 set. It could be shown that the average error decreased. Hence, the core-core term in the original MNDO method was replaced with a simple diatomic parameter function. A few other minor modifications were made [137]. Finally, parameters for the whole of the main group and some additions, 42 elements in all, were then optimized simultaneously. Two other approaches had been considered, but these were not complete (PM4) or not published (PM5), so the new method was termed PM6 [33].

2.3.4 DFTB

Density functional tight binding can be derived from a Taylor series expansion of the Kohn-Shame density functional total energy [141] around a properly chosen reference density $\rho(r)$. Rather than minimizing the energy and finding the electron density $\rho(r)$, a reference density ρ^0 is assumed and perturbed by some density fluctuation,

$$\rho(r) = \rho^0(r) + \delta\rho(r) \tag{2.16}$$

and the exchange-correlation energy functional is then expanded in a Taylor series up to the third order and the total energy can be written as

$$\begin{aligned} E = & \frac{1}{2}\sum_{ab}\frac{Z_a Z_b}{R_{ab}} - \frac{1}{2}\int\int\frac{\rho^0(r)\rho^0(r')}{|r-r'|}drdr' - \int V^{XC}[\rho^0]\rho^0(r)dr + E^{XC}[\rho^0] \\ & + \sum_i n_i\left\langle\psi_i\middle|\hat{H}^0\middle|\psi_i\right\rangle + \frac{1}{2}\int\int\left(\frac{1}{|r-r'|} + \left.\frac{\delta^2 E^{XC}[\rho]}{\delta\rho(r)\delta\rho(r')}\right|_{\rho^0}\right)\delta\rho(r)\delta\rho(r')drdr' \\ & + \frac{1}{6}\int\int\int\left.\frac{\delta^3 E^{EX}[\rho]}{\delta\rho(r)\delta\rho(r')\delta\rho(r'')}\right|_{\rho^0}\delta\rho(r)\delta\rho(r')\delta\rho(r'')drdr'dr'' \\ & = E^0[\rho_0] + E^1[\rho_0,\delta\rho] + E^2[\rho_0,(\delta\rho)^2] + E^3[\rho_0,(\delta\rho)^3] \end{aligned} \tag{2.17}$$

Depending on the inclusion of terms from this expansion, different models appear, which can be viewed as successively build on top of each other. The first of this models was the first-order non-self-consistent DFTB1 [142, 143], followed by the self consistent second-order DFTB2 (also called SCC-DFTB) [120] and the third-order extension DFTB3 [144, 86, 145, 146].

2.3.4.1 DFTB1

The DFTB1 method uses the first two contribution of equation (2.17), $E^0[\rho_0]$ and $E^1[\rho_0,\delta\rho]$. Also in DFTB, a linear combination of atomic orbitals (LCAO) ansatz of the KS orbitals are used:

$$\psi_i = \sum_\mu c_{\mu i}\phi_\mu \tag{2.18}$$

Where the AOs are obtained from DFT calculations of the corresponding atoms. Again, the basis is restricted to the valence shell of the atoms. Within the LCAO approach we get a general eigenvalue problem of the form

$$\sum_\nu c_{\nu i}(H^0_{\mu\nu} - \epsilon_i S_{\mu\nu}) = 0 \text{ with } \nu \in b \text{ and } \mu \in a, \tag{2.19}$$

with the Hamiltonian matrix elements $H^0_{\mu\nu}$ and the overlap matrix elements $S_{\mu\nu}$. Considering the orthogonality constraint and writing the effective KS potential V_{eff} as a superposition of atomic-like potentials ($V_{eff} \approx \sum_c V^c_{eff}$), one can write

$$(\phi_\mu|V^c_{eff}|\phi_\nu) \rightarrow (\phi_\mu|\left[V^c_{eff} - \sum_{K_c}|\phi^c_K)\epsilon_{K_c}(\phi^c_K|\right]|\phi_\nu) \text{ with} \nu \in b \text{and } \mu \in a, \tag{2.20}$$

where ϵ_{K_c} is the energy of a core state K at center c. The potential can also be viewed as pseudo-potential with its core correction term. This term only appears in the three-center terms of the Hamiltonian matrix elements ($a \neq b \neq c$) and in the 'diagonal terms' $\nu, \mu \in a$ with $c \neq a$, whereas the 'full' potential appears in all the other terms

$$H^0_{\mu\nu} = \left(\phi_\mu| -\frac{1}{2}\Delta^2 + \sum_j V^c_{eff}|\phi_\nu\right) - \sum_c\sum_{K_c}(\phi_\mu|\phi^c_K)\epsilon_{K_c}(\phi^c_K|\phi_\nu) \;\; \mu,\nu \notin \{c\} \tag{2.21}$$

Hence, the neglect of the three-centre terms and pseudo-potential contributions give a two-centre approximation for the Hamilton matrix elements

$$H^0_{\mu\nu} = (\phi_\mu| -\frac{1}{2}\Delta^2 + V^a_{eff} + V^b_{eff}|\phi_\nu), \;\; \mu \in \{a\}, \;\; \nu \in \{b\}. \tag{2.22}$$

The approximations lead to the same structure as in tight binding or extended Hückel method; Nevertheless, all matrix elements are calculated within DFT, since the AO basis set ϕ_μ is computed by solving the DFT-KS equations for atoms. Pure AOs would be too diffuse for a minimal AO basis set; therefore, the atomic KS equations are usually solved applying an additional (harmonic) potential to the atomic KS equations, which in turn leads to 'compressed' AOs, as introduced by Eschring [147]

$$\left[-\frac{1}{2}\Delta^2 + v^{eff}[\rho^{atom}] + \left(\frac{r}{r^0}\right)^2\right]\phi_\mu = \epsilon_\mu\phi_\mu. \tag{2.23}$$

The confinement radius is taken to be roughly two times the covalent radius of the atom according to [143]. The electron density of the neutral atoms a are contained in the potentials $V^a_{eff} = V_{eff}(\rho_a)$. This density is usually also determined from atomic KS equations as well as the AO basis set. Nevertheless, different confinement radius can be chosen for these initial densities ρ_a and therefore two parameters per element have to be determined, the confinement radius for the AO basis r_0 and the the confinement radius for the initial density, r^d_0.
Now, with initial density and AO basis determined, the KS equations can be solved (equation (2.19)) leading to the energy (n_i:occupation number of KS orbital i)

$$E^1 = \sum_i n_i \sum_{\mu\nu} c^i_\mu c^i_\nu H^0_{\mu\nu} = \sum_i n_i\epsilon_i \tag{2.24}$$

This is the sum of the occupied KS energies or the electronic energy of the DFTB method. Additionally, to get the total energy, the E^0 term has to be approximated. $E^0[\rho_0]$ consists

of the DFT 'double counting' contributions in the first line of equation (2.17) and only depends on the reference density ρ_0. Resulting in a term that is not dependent on the specific chemical environment and therefore can be determined for a 'reference system' and then applied to other molecules in different environments and thus resulting in transferable parameters. Hence, $E^0[\rho_0]$ is approximated by a sum of pair potentials called repulsive energy term [119]

$$E^0[\rho_0] \approx E_{rep} = \frac{1}{2}\sum_{ab} V_{ab}^{rep}, \tag{2.25}$$

which can be determined by DFT calculations [143] or fitted to empirical data [148]. The expression for the total energy then results as

$$E^{DFTB1} = \sum_i n_i \epsilon_i + \frac{1}{2}\sum_{ab} V_{ab}^{rep} \tag{2.26}$$

Thus, $H_{\mu\nu}$ and $S_{\mu\nu}$ are computed once and stored and can be read in from tables. Further, since DFTB1 is a non-self-consistent TB method the KS equations are solved only once and resulting therefore in a speed-up of 5-10 compared to DFTB2 or DFTB3. Of course limitations are that DFTB1 is only suitable for systems with small charge transfer between the atoms. This is in particular the case for homo-nuclear systems or systems with atoms of similar electro-negativity. Hydrocarbons are systems where DFTB1 is also very well suited. For systems exhibiting a delicate charge balance, higher order terms have to be considered [144].

2.3.4.2 DFTB2 and DFTB3

The extensions DFTB2 and DFTB3 approximate the E^2 and E^3 terms in equation (2.17) respectively. The first assumption is, that the density fluctuations can be written as a superposition of atomic contributions

$$\delta\rho = \sum_a \delta\rho_a \tag{2.27}$$

and for the atomic-like density fluctuations a multi-pole expansion is used where only the monopole terms are kept

$$\delta\rho_a = \Delta q_a F_{00}^a Y_{00} \tag{2.28}$$

Practically, the second-order integral (third line in equation (2.17)) is taken to be an exponentially decaying charge density

$$\delta\rho_a \approx \Delta q_a \frac{\tau_a^3}{8\pi} e^{-\tau_a|r-R_a|} \tag{2.29}$$

and the Hartree integral is analytically evaluated. Resulting in an analytical function γ_{ab}, and the second-order term finally reads (for $a \neq b$)

$$E^2(\tau_a, \tau_b, R_{ab}) = \frac{1}{2}\sum_{ab} \Delta q_a \Delta q_b \gamma_{ab}(\tau_a, \tau_b, R_{ab}) \tag{2.30}$$

$$= \frac{1}{2}\sum_{ab} \delta q_a \delta q_b \int\int' \left(\frac{1}{|r-r'|} + \frac{\delta^2 E^{XC}[\rho]}{\delta\rho(r)\delta\rho(r')}\bigg|_{\rho^0} \right) F^a_{00} F^b_{00} Y^2_{00} dr dr' \tag{2.31}$$

The Hartree term therefore describes the interaction of the charge density fluctuation $\delta\rho_a$ and $\delta\rho_b$. For large distances, $R_{ab} = |r - r'| \to \infty$ the XC terms vanish and the integral describes the coulomb-interaction of two spherical normalized charge densities, which is basically $\frac{1}{R_{ab}}$ and hence $E^2 \approx \frac{1}{2}\sum_{ab} \frac{\delta q_a \delta q_b}{R_{ab}}$. On the other hand, for vanishing inter-atomic distance $R_{ab} = |r - r'| \to 0$, the integral describes the electron-electron interaction on atom a $E^2 \approx \frac{1}{2}\frac{\delta^2 E_a}{\delta^2 q_a} = U_a$ where U_a is know as the Hubbard parameter or the chemical hardness. It describes how much the energy changes upon adding or removing electrons. From the functional form of γ_{ab}, one finds that the Slater exponent of an atom a is related to the Hubbard parameter as

$$\tau_a = \frac{16}{5} U_a, \tag{2.32}$$

stating that the width of the atomic charge density is inversely proportional to its chemical hardness. This relation is intuitive in that more diffuse atoms have a smaller chemical hardness. Since γ_{ab} encodes a constant proportionality, some problems arise because this relations seems not to be valid across the periodic table. The deviation is most pronounced between hydrogen and the first-row elements. To correct for this, a modified γ^h_{ab} has been proposed [144, 85, 146] and a further explanation will follow in chapter 3.7.1.

For E^3, the same approximation are introduced as for E^2. The third-order terms describe the change of the chemical hardness of an atom with its charge state [10] and hence a new parameter is introduced, the chemical hardness derivative U^d_a. Initially, only the diagonal terms were included in the third order terms [144, 85, 146] . The U^d_a parameters can be computed from DFT or fitted. As a result Γ_{ab} emerges as the derivative of the γ-function with respect to charge by introducing the Hubbard derivative parameter. Hence, the third-order terms can be viewed as charge dependence capturing some problems of the second-order formalism.

With all these approximations, the SCC-DFTB total energy with the third-order expansion is given by

$$E^{DFTB3} = \sum_{iab}\sum_{\mu\in a}\sum_{\nu\in b} n_i c c_{\nu_i} H^0_{\mu\nu} + \frac{1}{2}\sum_{ab}\Delta q_a \Delta q_b \gamma^h_{ab} + \frac{1}{3}\sum_{ab}\Delta q_a^2 \Delta q_b \Gamma_{ab} + \frac{1}{2}\sum_{ab} V^{rep}_{ab} \tag{2.33}$$

The derivative of the expression with respect to the molecular orbital coefficients leads to the corresponding KS equations

$$\sum_{\nu} c_{\nu i}(H_{\mu\nu-\epsilon_i S_{\mu\nu}}) = 0 \text{ with } \nu \in b \text{ and } \forall a, \mu \in a \tag{2.34}$$

and

$$H_{\mu\nu} = H^0_{\mu\nu} + S_{\mu\nu}\sum_c \Delta q_c \left(\frac{1}{2}(\gamma^h_{ac} + \gamma^h_{bc}) + \frac{1}{3}(\Delta q_a \Gamma_{ac} + \Delta q_b \Gamma_{bc}) + \frac{\Delta q_c}{6}(\Gamma_{ca} + \Gamma_{cb})\right) \tag{2.35}$$

where $S_{\mu\nu}$ is the overlap matrix. The Hamilton matrix elements depend on the Mulliken charges $q_a(\Delta q_a = q_a - Z_a)$ which in turn depend on the molecular orbital coefficients $c_{\mu i}$, and hence these equations have to be solved self-consistently.
Closing, it has to be noted that the E^3 term consists of diagonal and off-diagonal parts. Originally, only the diagonal terms have been included and to date the current implementation of CP2K [149] also considers only the diagonal contributions to the third-order terms. Hence, in principle the term DFTB3-diag should be employed but throughout this manuscript the shorter form DFTB3 is used consequently even though the diagonal terms are not included.

3 Applications

In this chapter the implementation details of the multi-objective optimization algorithm are given as well as the mechanisms of how the multi-dimensional handling of the Pareto set and the selection (decision making) thereof was carried out. Additionally, the entire multi-objective set-up is summarized and the use of the selected objective functions shortly discussed. The importance of the reference data is highlighted and the road towards our final data selection elaborated. Finally, the results for the $PM6_{optim}$ and $DFTB3_{optim}$ simulations of liquid water at ambient conditions are compared with the standard parameters, classical force field approaches and ab-inito calculations.
For the oxygen radial distribution function at ambient bulk water condition substantial improvements for both semi-empirical methods could be achieved. Enhancement can also be reported for the final ambient densities for water calculated with $PM6_{optim}$ and $DFTB3_{optim}$. The obtained densities for the optimized semi-empirical methods are even comparable to ab-inito calculations with dispersion correction (BLYP-D3) and therefore render a more realistic picture of modelled water with the newly employed parameter set. Analysis of the slab simulation, the performance of the semi-empirical methods appear to be comparable to the classical force-fields when it comes to surface tension. With regard to the self-diffusion coefficient values for both semi-empirical methodss, it could be found that they are in better agreement than ab initio values and force field methods.
Thus, the finding demonstrates the importance of the optimization procedure and the potential of many models or methods that could be exploited with a systematic re-parametrization. An other important issue that should be emphasized is that before correcting methods for deficiencies (like PM6-DH [150] corrects for dispersion and hydrogen bonding errors), it should be made sure that the short comings are not based on improper parameter sets. Thus, our main contribution in this chapter is to show that the NDDO model (i.e PM6) is indeed capable of reproducing many difficult properties, which are encountered employing water simulations.

3.1 Introduction

All sorts of theoretical and technical problems arise in the area of parameter optimization and hence it constitutes an independent field within applied mathematics. Employing the theoretical framework to real (physical) problem imposes an extra layer of difficulty to the already complicated process. Parameter fitting is a key process in computational chemistry even for ab-initio methods since the quantum-mechanical systems can not be solved analytically. However, there are three interrelated problems associated with parameter optimization in this field which are all incomplete:

1. computational power
2. historical
3. systematic

In the beginning of computational chemistry and with the development of computer technology starting 1940 the *computational* power was limited. Thus, for parametrization purposes only limited sets of reference data or limited sets of parameter itself could be optimized and as a result thereof, incomplete optimization results were obtained (only certain parameter were optimized where others were kept fixed). Despite of this caveat, results were promising, with few known defects arising from the methodology. The success of the obtained parameter sets gave rise to a resulting problem. The obtained parameter were kept constant over time (*historical legacy*) even when fundamental changes to methods were proposed. Certainly, one should re-optimize the parameter set every time changes to the model are adopted. A prominent example was the development of PM3 parameter sets where first initial elements were fitted and then kept fixed for the addition of new elements.
Finally, there is no consensus about how to carry out parameter optimization in computational chemistry. For the development of semi-empirical methods usually reference enthalpies, geometries, ionization potentials, dipoles amongst others are used. Thus, all parametrizations employ different scalarization and selections and thus lack a *systematic* approach. This raises challenges to reproducibility of optimized parameter sets and hence a much cleaner and simpler approach such as 'real' multi-objective optimization should be carried out. Therefore, the main goal of this manuscript is to advocate a more integral view on parameter optimization in computational chemistry.

3.2 Mechanisms

Here we review both, the implementation of the multi-objective optimization algorithm as well as the mechanisms of how to deal with a Pareto set in a multi-objective environment. Thus, shortly some visualizations and selection options are presented and then applied to

the parameter optimization problem of the two semi-empirical methods described earlier on.

3.2.1 Implementation

In chapter 1 we outlined the schema of multi-objective optimization with the use of covariance matrix adaptation. The entire implementation in CP2K [149] is based on a simple master-worker design (Figure 3.1) accomplished with the standardized and portable message-passing system (MPI) available for Fortran and perfectly suited for high performance computing. The entire code is available as a fork of the official CP2K package on github [151]. Some minor changes to the official code were made such as energy convergence flags since those are crucial to ensure an unobstructed optimization process. But mostly the multi-objective code is autonomous and therefore little changes to the original CP2K program is necessary since its mostly based on the particle swarm framework developed by Ole Schütt [149].

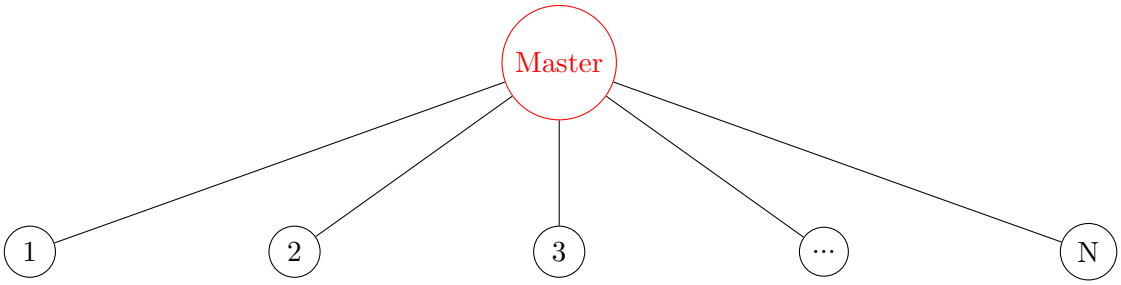


Figure 3.1 – Master and worker scheme as implemented in CP2K to carry out the multi-objective optimization.

Theoretically, for each individual in the population (from the MO-CMA-ES) which is represented as a 3-tuple $(x_k, s_k, f^m(x_k))$, one needs to perform m function evaluations depending on the number of objectives m quantified. Usually in practice though, only one quantum chemistry calculation has to be performed to achieve all the desired outcomes and hence all the m objectives at once. Carrying out a geometry optimization results in properties such as the geometry, energy and atomic charges which typically are compared to reference values and taken as objectives. This leaves us with one calculation per individual and hence based on a direct mapping of individual and worker distributes the task over all dynamically available worker. The number of workers can be chosen resource depended and theoretically set up to the hardware limit of the computing resource. This flexible scheme allows a perfect scaling for high performance computation since by increasing the size of the population in an multi-objective optimization run one can simply increase the number of workers to match the desired time per iteration. This is illustrated in table 3.1, where the necessary tasks per multi-objective optimization iteration are given. In that given example n geometry optimizations (based on the

reference data) have to be carried out for k individuals. Therefore, the master node distributes this table (the parameter set x_k respectively) to the available worker and receives in return the objectives $f^m(x_k)$. It is very difficult to employ a perfect use of the

individual	structure 1	structure 2	structure $\cdots$	structure n
1	$f^m(x_1)$	$f^m(x_1)$	$f^m(x_1)$	$f^m(x_1)$
2	$f^m(x_2)$	$f^m(x_2)$	$f^m(x_2)$	$f^m(x_2)$
$\cdots$	$f^m(x_{...})$	$f^m(x_{...})$	$f^m(x_{...})$	$f^m(x_{...})$
k	$f^m(x_k)$	$f^m(x_k)$	$f^m(x_k)$	$f^m(x_k)$

Table 3.1 – Internal data structure of the tasks to distribute amongst workers. For every structure 1 to n in the reference set properties have to be calculated f^m for all x_k parameter sets ($n \times k$ matrix)

available computing resources since the structures may differ subsequently and hence the required computational time for geometry optimization runs vary enormous between different structures. This heterogeneous setup subsequently leads to idle time for some workers that only could be resolved by grouping workers to flexible work groups to achieve a balanced work load and hence a more efficient usage of the computing resources. But predicting necessary computational resource for a given structure and parameter set is almost impossible and hence the simpler one process per worker ratio (i.e one MPI process per geometry optimization) was maintained.
However, some features were implemented to speed up and avoid unnecessary computations. On one hand, if for one parameter set x_k the initial energy of a structure can not be converged then all the other structures are skipped and no calculations are carried out for that set. On the other hand, if during the calculations of the geometry minimum the structure does not achieve geometry convergence, flags are passed such that the remaining structures are not processed. Thus, undesired parameter sets are intercepted as early as possible to speed up the overall optimization iteration.

3.2.2 Visualization of n-dimensional Pareto front

As described in Chapter 1, obtaining the Pareto front is not the last step in the decision making process. One or more candidate solutions have to be selected from the performance space for further inspection. It is widely accepted that visualization tools are valuable and provide decision-makers with a meaningful method to analyse the Pareto set and select good solutions. For bi-objective problems it is usually straightforward to make an graphical analysis of the Pareto set points possibilities, but this becomes more difficult for higher dimensions. The most common are summarized in Table 3.2.

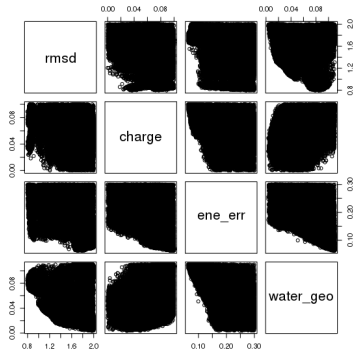
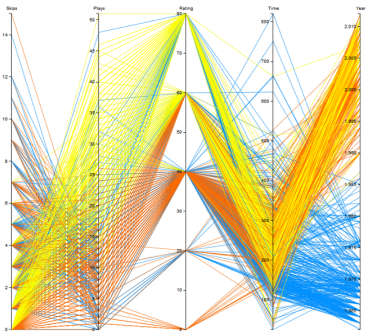
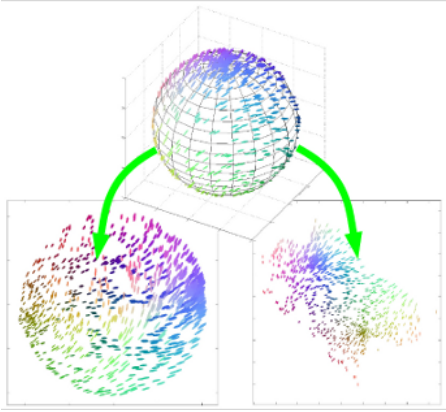
Method	Illustration
Scatter diagrams: The visualization consists of an array of scatter diagrams arranged in the form of an $n \times n$ matrix. Each dimension of the original data defines one row and column of the matrix. The complexity of the representation increases notably with the dimension but all pair-wise Pareto fronts are shown simultaneously and correlations are easily detected.	
Parallel coordinates: A multidimensional point is plotted in a two-dimensional graph. Each dimension of original data is translated to an x-coordinate in the two-dimensional plot. This is a very compact way of presenting multidimensional information, but with large sets of data it loses clarity and the analysis becomes difficult to perform	
Dimensional reduction: Another approach would be to apply a dimensional reduction $\mathbb{R}^n \rightarrow \mathbb{R}^3$ so that the Pareto front can be displayed in usual manner in $\mathbb{R}^3$ or from $\mathbb{R}^3$ reduced to a much simpler $\mathbb{R}^2$ plot as shown. In our specific case dealing with a 4th dimensional decision space the last two dimension were merged with the feature scaling scalarization. For that, the objectives f_3, f_4 were rescaled to be within range $[0, 1]$. The general formula is: $$x' = \frac{x - min(x)}{max(x) - min(x)}$$ where x is an original function value in the given objective hence x' is the normalized value. Consequently, after normalizing objectives f_3 and f_4 the resulting $f_{cons} = f_3' + f_4'$ is constructed. Hence, a normal 3-dimensional plot can be obtained with the caveat that the constructed dimension is somewhat difficult to interpret.	

Table 3.2 – Visualization methods of n-dimensional Pareto front

3.2.3 Selection from n-dimensional Pareto sets

Our strategy of selection process started with the visual inspection of the pairwise scatter diagrams. The plots can already indicate what regions of the multi-dimensional space should be investigated in more detail. Usually, certain ranges can be excluded to narrow down the potential solutions of interest. After a pre-selection was carried out, we picked k-solutions by the k-means clustering approach of the remaining Pareto set. Given a set of observations $(x_1, x_2, \cdots, x_n)$, where each observation is a d-dimensional real vector, k-means clustering aims to partition the n observations into $k(\leq n)$ sets $S = S_1, S_2, \cdots, S_k$ so as to minimize the within-cluster sum of squares (WCSS). In other words, its objective is to find:

$$\underset{S}{\operatorname{argmin}} \sum_{i=1}^{k} \sum_{x \in S_i} \|x - \mu_i\|^2 \tag{3.1}$$

where μ_i is the mean of points in S_i. A simple illustration is given in Table 3.3 where every step is listed to describe the algorithm. In the current implementation, the k-means

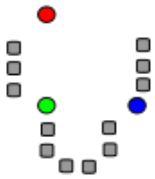
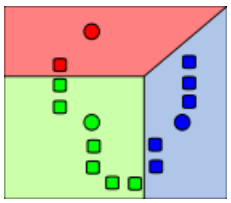
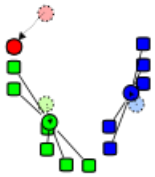
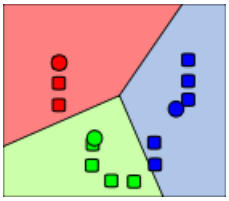
			
1. k initial "means" (in this case $k = 3$) are randomly generated within the data domain (shown in color).	2. k clusters are created by associating every observation with the nearest mean. The partitions here represent the Voronoi diagram generated by the means.	3. The centroid of each of the k clusters becomes the new mean.	4. Steps 2 and 3 are repeated until convergence has been reached

Table 3.3 – Demonstration of the standard algorithm for the k-mean clustering in 4 steps

clustering is replaced by the k-medians clustering. It is a variation of the k-clustering method where instead of calculating the mean for each cluster to determine its centroid, one instead calculates the median. This has the effect of minimizing error over all clusters with respect to the 1-norm distance metric, as opposed to the square of the 2-norm distance metric [152].
This relates directly to the k-median problem which is the problem of finding k centres such that the clusters formed by them are the most compact. Formally, given a set of data points x, the k centres c_i are to be chosen so as to minimize the sum of the distances

from each x to the nearest c_i. Since we want to obtain k representative of our Pareto set, this is identical of finding the k centres as described above.

3.3 Multi-objective setup

The strategy parameter are chosen as default values, according to Table 1.1 from the first chapter. It has to be noted that the entire optimization process is carried out in scaled vector space. Thus, a hypercube scaling is employed for the parameters where the resulting values are within $[0 \cdots 1]^n$. This transformation allows to initialize the covariance matrix adaptation algorithm with basic initial parameters according to subsection 1.3.5.2. However, upper and lower limits of the variable space have to be set for this transformation and hence the PM6 parameters were allowed to be within a bound of ±10%. A small value was chosen on purpose for the parameter band since this might guarantee a better transferability to the old values and hence makes it easier to compare with standard PM6 calculations. It is best to maintain the parameter set where possible since the standard PM6 parameter value have proven to perform reasonable in a lot of applications [153, 154, 155, 156, 157] but not for bulk water [158]. As for the DFTB3 optimization the boundaries were set to bound_{low} = [0.5, 0.0, 2.5] and bound_{upper} = [3.5, 1.0, 5.0] from visual inspection of the γ-function and the proposed correction respectively. Initial values for the PM6 are straight forward since there are standard values available [33]. For the DFTB3 initial values were set to $a = 1.0$, $b = 0.5$, $c = 3.0$ again from analytical formal inspection. The initial step size σ was set small (0.01) such that an initial local exploration is enforced. The population size was chosen to be 800 for the PM6 optimization since the decision space is huge ($\mathbb{R}^{27}$) and hence a large population is needed. Although for the DFTB3 the decision space is much smaller, the same population size was kept for practical reasons. A short summary of the most important values is given in Table 3.4.

3.4 Objective Functions

The parametrization of the original MNDO method focused on ground-state properties, mainly heats of formation, geometries, ionization potentials and dipole moments. AM1 and PM3 followed the same philosophy but were more extensive in respect to reference data sets. Enthalpy f_1 and geometry f_2 are basic requirements and therefore have to be included as objectives. The ionization potentials were neglected since we are carrying out a re-optimization and keeping the parameters in boundaries and hence should in principle grantee that the ionization potential is not affected largely. The water monomer dipole is not chosen as objective function since it is a compound property of its geometry f_3 and partial atomic charges f_4. Hence, we decided to include an error on atomic charge for oxygen f_4 as last objective. The geometric error was calculated according to the root-mean-square-deviation explained below and is a new entity since usually the bond lengths error or angle deviation are used as measure for geometric accuracy. The relative

Description	Parameter	Value
Step size damping	d	$1+\frac{n}{2}$
Target success probability	p_{succ}^{target}	$\frac{1}{5+\sqrt{\frac{1}{2}}}$
Success rate averaging parameter	c_p	$\frac{p_{succ}^{target}}{2+p_{succ}^{target}}$
Covariance matrix adaptation	c_c	$\frac{2}{n+2}$
Covariance learning rate	c_{cov}	$\frac{2}{n^2+6}$
Boundaries	$\text{bound}_{lower}^{upper}$	±10% of starting value for PM6
Covariance matrix	C	Identity matrix I
Step size	σ	0.01
Population size	λ	800

Table 3.4 – Multi-objective set-up values for the most important parameters and short description thereof.

enthalpy error for all compounds in the reference set was obtained as described in the subsection 3.4.2 and follow a more traditional approach. The error from the Mulliken charges are employed according to subsection 3.4.3 and a more detailed description on the selection of reference data is given in 3.5 further on.
In principle one could also employ performance related objectives such as minimize SCF cycles or faster convergence of geometry optimization. This would open up new fields in basis set optimization techniques for ab intio parametrization and one could re-optimize for accuracy and speed before carrying out time consuming geometry or molecular dynamics simulations.

3.4.1 Root-Mean-Square-Deviation of atomic positions (RMSD)

The root-mean-square deviation (RMSD) is the measure of the average distance between the atoms of the superposition of the reference geometry and the optimized structure. Coutsias, et al. [159] presented a simple derivation, based on quaternions, for the optimal solid body transformation (rotation-translation) that minimizes the RMSD between two sets of vectors [159].

$$\epsilon_{RMSD}^{i} = \sqrt{\frac{1}{M}\sum_{j=1}^{M}\delta_j^2} \tag{3.2}$$

where δ is the distance between N pairs of equivalent atoms. Usually a rigid superposition which minimized the RMSD is performed, and this minimum is returned. Given two sets of n points v and w, the RMSD is defined as follows:

$$\epsilon_{RMSD} = \sqrt{\frac{1}{M}\sum_{j=1}^{M} \| v_j - w_j \|^2} = \sqrt{\frac{1}{N}\sum_{j=1}^{2} \left((v_{jx} - w_{jx})^2 + (v_{jy} - w_{jy})^2 + (v_{jz} - w_{jz})^2 \right)} \quad (3.3)$$

Commonly the RMSD is expressed in length units usually in Angstroms [Å] for our purpose.
The main idea behind using the RMSD is the fact that the reference geometries are in equilibrium. A structure is in equilibrium when all forces or moments acting upon it are balanced. This means that each and every force acting upon a body, or part of the body, is resisted by either another equal and opposite force or set of forces whose net result is zero. Thus, the net force acting on each atom of an equilibrium structure must therefore be negligible. Consequently, instead of using a force matching approach where the initial forces of a geometry optimization run should vanish, we decided to include the "dynamical" relaxation and hence compare the reference geometry with the optimized structure. It is clear that a small error in RMSD correlates with a small error in the forces in the geometry optimization process. Therefore using the RMSD is a far more integrated approach. The limiting factor of using the RMSD as measure is that the reference structures need to be in equilibrium state otherwise eventually something like force matching approaches must be applied.
Finally the sum is taken for all the RMSD error of the structures in the reference set and thus the value for the error of the geometries reads:

$$\epsilon_{RMSD} = \sum_{i=1}^{N} \sqrt{\frac{1}{M}\sum_{j=1}^{M} \delta_j^2} \quad (3.4)$$

3.4.2 Enthalpy Error

The relative error was calculated to quantify the offset in heat of formation. The heat of formation ΔH_f is a measure for relative stability and calculated in gas phase usually at 298 K. There would be no point in employing absolute energies for parametrization since they can not directly be compared. On the other hand heat of formations can directly be investigated since they are relative to the compound basis energy. To account for the different system sizes and enthalpies in the reference data, the absolute error $|\Delta H^k_{ref} - \Delta H^k|$ is divided by the reference enthalpy ΔH^k_{ref} resulting in a relative deviation per structure. Consequently, all the contributions of the errors in enthalpy are summed

to the final value of $\epsilon_{enthalpy}$. We can express the energy error as

$$\epsilon_{enthalpy} = \sum_{k=1}^{N} \frac{|\Delta H_{ref}^{k} - \Delta H^{k}|}{\Delta H_{ref}^{k}} = \sum_{k=1}^{N} \frac{\Delta\Delta H^{k}}{\Delta H_{ref}^{k}} \tag{3.5}$$

where the sum is taken over all N structures in the reference set. Through the manuscript the error in enthalpy is usually given in kcal/mol when not further stated. Thus one should keep in mind that the reported values are always *relative* enthalpy errors.

3.4.3 Charge Error

The calculation of effective atomic charges plays an important role in the application of quantum mechanical calculations to molecular systems [7, 9, 160]. This is despite all conceptual problems connected to dividing up the overall molecular charge density in atomic contributions, and all practical problems connected to finding a convenient and robust algorithm applicable to a wide range of systems [161]. Partial atomic charges play a crucial role in reactivity and account for the resulting dipole of a molecule. Although the Mulliken population analysis [7, 9] has some deficiencies, it is a well established and widely used method to divide the charge density. Hence, atomic partial charges are calculated and compared with the reference atomic charge. The ground-state structure of the water molecule has been studied repeatedly [162, 163], and we will use here the structure suggested by Benedict and co-workers [162] with $r_{OH} = 95.72°$ and $a_{HOH} = 104.52°$. The gas phase molecular dipole moment of water has been measured to $1.855D$ [164]. Assuming the structure of water to be rigid at its experimental geometry, this corresponds to effective atomic charges of $q(O) = -0.66e$ and $q(H) = +0.33e$. Therefore, the charge error is defined as the difference in oxygen atomic Mulliken charges to the reference charge

$$\epsilon_{charge} = q(O)_{ref} - q(O)_{calc} = \Delta q(O) \tag{3.6}$$

Since the dipole is an entity consequently resulting from geometry and charge, it makes more sense to use those attributes because one could obtain a correct dipole with wrong geometry and charges.

3.5 Reference Data

The selected reference data should be representative of the target system and also be within the theoretical scope of the underlying model. Hence the reference data should be able to be reproduced within the frame-set available. For instance, the water27 (description in 3.5.1) would not be a reference database [165] eligible for classical force field methods because it contains protonated water clusters which can not be described with the typical force field approaches.

Since the goal was to re-parametrize for bulk water simulation, some available compilations of water clusters were considered. It should be mentioned that the selection of the reference data is a very delicate undertaking. In principle, different reference data should be systematically tested together with the complete optimization routine and the underlying method. Thus, for small reference sets multi-objective optimization might provide some insights. For example multi-objective optimization could be used to investigate the geometries for different water cluster sizes starting from the monomer $n = 1$ where the parameters of the model (i.e PM6 parameters) are only optimized with respect to the monomer geometry as objective (f_1). Then, systematically n (cluster size) is increased and more objectives f_n are added. But as elaborated in the first chapter, adding more objectives render the entire optimization process unpractical and hence we decided to consider well established sets from the literature as described in the following sections:

3.5.1 Water27

Water27 reference data [165] is a compilation that consists of 27 cluster binding energies extrapolated to the CBS limit of the MP2 and CCSD(T) theory. This includes a set of 14 neutral water clusters, 5 hydronium ion clusters, 7 hydroxide ion clusters, and 1 auto-ionized water cluster. The structures are shown in Figure 3.2, and the Cartesian coordinates are obtained after optimization at the B3LYP/6-311+G(2d,2p) level. Although the compiled set of Bryantsev et al [165] seems like a valid choice it leaves open the room for speculation. It is not clear which combination of clusters represents the best model for bulk water. Presumably, the inclusion of charged clusters is not necessary for a valid liquid bulk water description and hence a smaller set should be used for practical reasons. For every optimization process it is always desirable to keep the reference system as small as possible since the arising calculations increase with the size of the reference set.

3.5.2 BEGDB water clusters

BEGDB (Benchmark Energy and Geometry DataBase) selection is a set of global and local minima of water cluster containing 2-10 waters [166]. The geometry (Figure 3.3) of each isomer was optimized on a RI-MP2/aug-cc-pVDZ basis. The binding energies were calculated by extrapolating to the complete basis set limit. It can be argued that the compilation by Temelso and Shields [166] contains more realistic features of bulk water systems than the water27 database and therefore should be preferred as reference system. However, the reference system should not be redundant to wit the same feature should only occur proportional to the target system. In that respect, the BEGDB water cluster set seems not to be ideal but since it is not clear which features represent the liquid bulk water system the best no systematic reduction can be made.

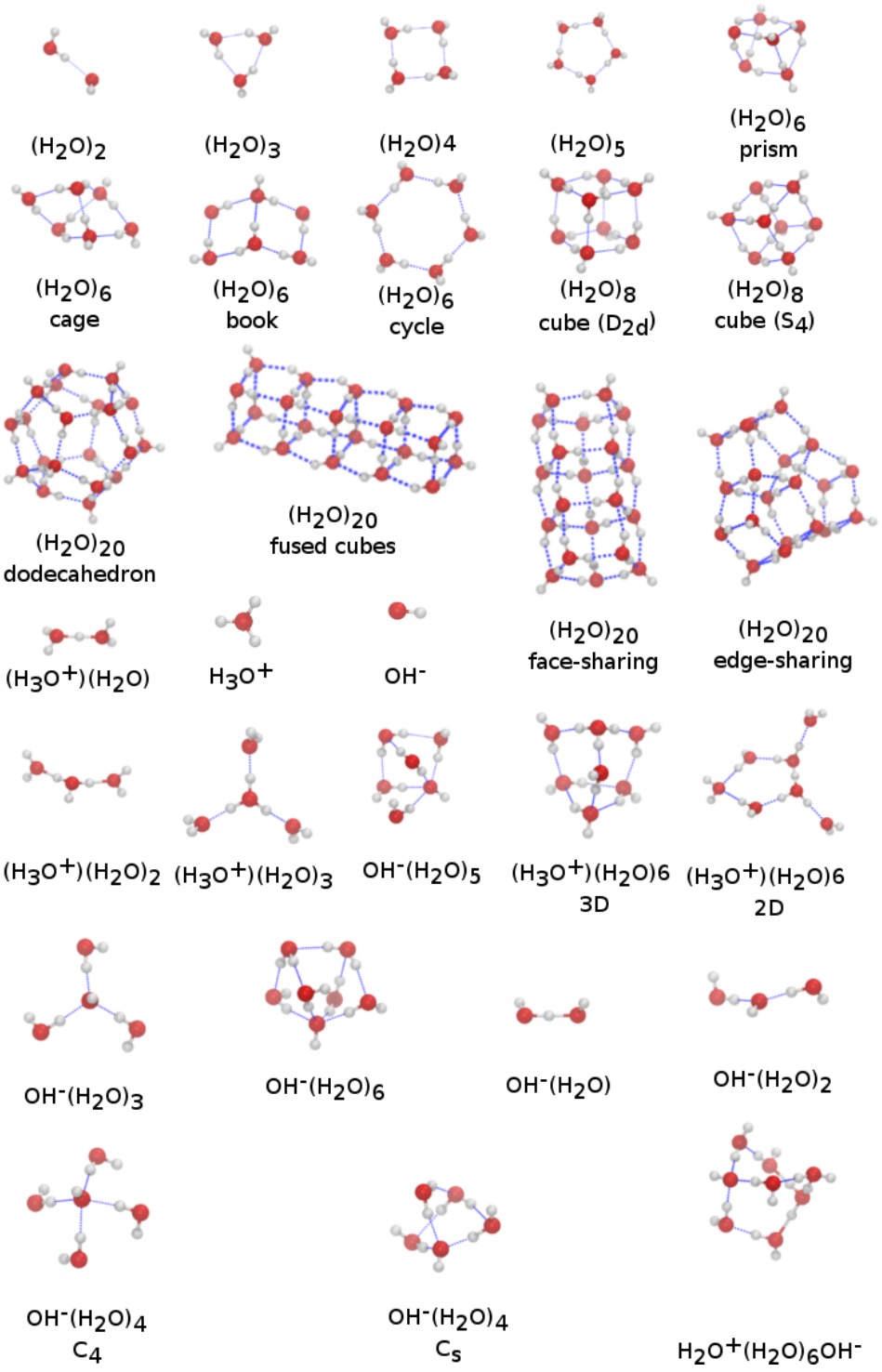


Figure 3.2 – Structures from Bryantsev [165] pure, protonated, and deprotonated water complexes further refered as water27 set.

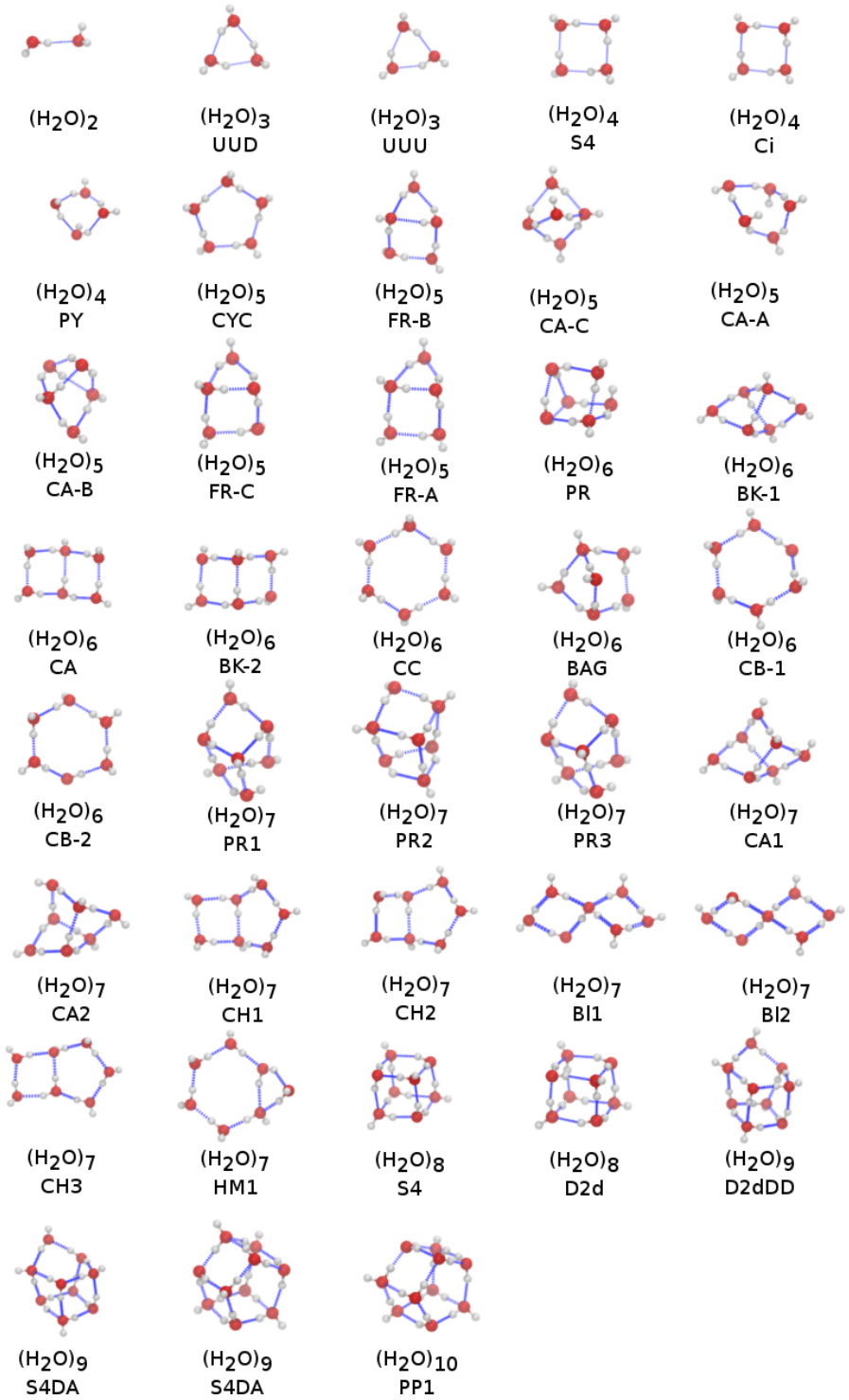


Figure 3.3 – RI-MP2/CBS low energy isomers of $(H_2O)_{n=2-10}$ sorted in order of increasing size and electronic energy and referred to as BEGDF water set [166]

3.5.3 QCE clusters

QCE (Quantum Cluster Equilibration) research employs the question "what can clusters tell us about the bulk?". The quantum cluster equilibrium was first published in 1998 by Frank Weinhold [12]. A concise definition of the QCE method comprises the essential idea of applying statistical mechanics to quantum chemically calculated clusters in order to gain insight into liquid and vapour phase. One helpful result of these calculations is the population composition of the different clusters considered for each phase point. Several results were obtained with the QCE models for water population analysis [13, 167, 168]. As example Figure 3.4 taken from Ralf Ludwig [168] is shown where the population

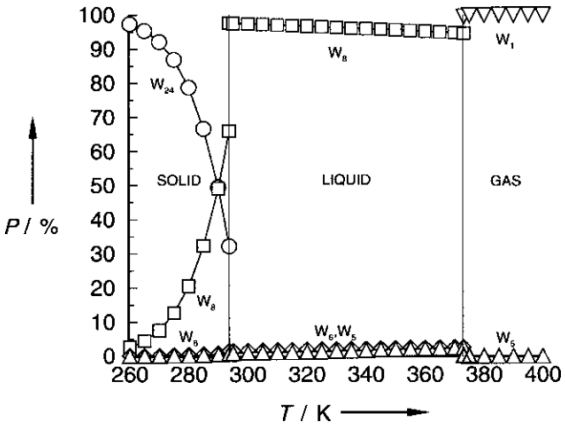


Figure 3.4 – Population analysis by Ludwig et al [168]. The population of certain clusters are shown for different phases. In this analysis the predominant cluster for liquid appeared to be w8 [168].

of different water clusters (in %) is shown for different temperature regimes and hence phases. A more recent study by Lehmann et al [13] advocates that $(H_2O)_9$ (s9) is predominant for the ambient conditions in the bulk as shown in Figure 3.5. In a more extensive analysis of different reference sets the final selection consisted of w1, w2, w3B, w5, w6, w8CUBE, s9, s11 as they are predominant for the liquid bulk water phase and hence the ideal candidate as reference system. As shortly described in [13], structures and energies were obtained employing MP2 (TZVPP) theory together with resolution of identity (RI) approximation [169]. The resulting reference values for the enthalpy are presented in Table 3.5.

3.6 Reparametrizing PM6

The main goal of the re-parametrizing process is twofold. On one hand, the multi-objective optimization scheme should be tested and hence a systematic approach established. As target we chose bulk water simulations since this system is not easy to reproduce [170] and only two elements (H and O) would be needed to be re-parametrized. Hence, on

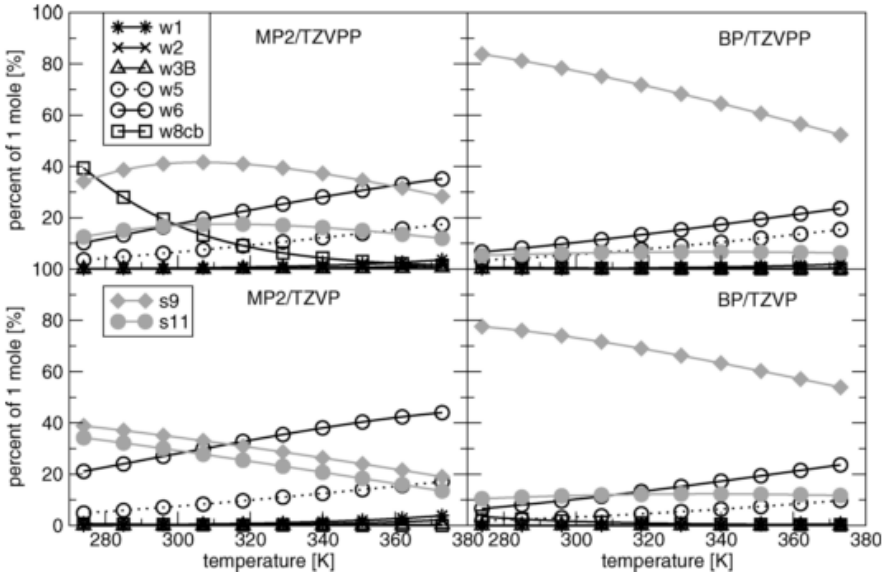


Figure 3.5 – Populations at the liquid-phase temperature range for different sets [13]

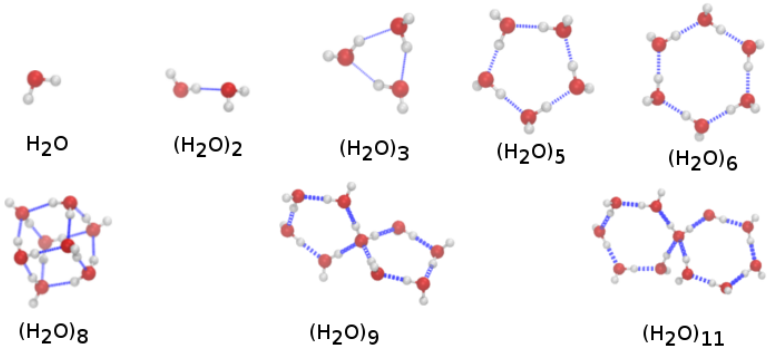


Figure 3.6 – The final selection of clusters of Lehmann et al [13] . The set includes strctures w1, w2, w3B, w5, w6, w8cube, s9, s11 with detailed selection process described in [13].

Structure	Reference Enthalpy [kcal/mol]
w1	-
w2	4.58891
w3B	14.006
w5	33.6759
w6	41.826
w8cube	67.1367
s9	67.3279
s11	81.9312

Table 3.5 – The reference enthalpy values in [kcal/mol] are shown for the final set [13].

the other hand the question whether the underlying theory of the PM6 method allows realistic description of bulk water should be answered.
In the course of the optimization process, an additional very important question is addressed, namely how the reference data should be chosen. In our particular case we sought to understand whether bulk water properties can be reproduced from water cluster reference data.

3.6.1 Parameters

As we have seen in chapter 2, all semi-empirical methods contain sets of parameters. Usually they only consider the valence s- and p-functions, which are taken as Slater type orbitals with corresponding exponents and different integrals according to the level of theory. A short description of the parameters for PM6 are given here. For the re-optimization of the default PM6 parameters there are 7 parameters for the hydrogen atom and 14 parameters for the Oxygen atom as well as an additional 6 parameters from the diatomic contributions. In total, hence we face a 27-dimensional search space of our optimization problem $\mathbb{R}^{27}$. In table 3.6 all occurring parameters in the PM6 method are listed. The one-center one-electron integrals have a value corresponding to the energy of

pm6 parameters for H and O	
One-center two electron repulsion integrals (eV)	$G_{ss}, G_{sp}, G_{pp}, G_{p2}, H_{sp},$
One-center one-electron integrals (eV)	U_{ss}, U_{pp}
Resonance (eV)	β_s, β_p
Slater exponent (a.u.)	ζ_s, ζ_p
Core exponent ($Å^{-1}$)	a,b,c
Diatomic ($Å^{-1}$)	$a_{HH}, x_{HH}, a_{OO}, x_{OO}, a_{OH}, x_{OH}$

Table 3.6 – A conceptual breakdown of the parameters in PM6 where the first three lines show the integrals followed by the Slater exponent and core-core repulsion parameters.

a single electron experiencing the nuclear charge (U_{ss}, U_{pp}) plus terms from the potential due to all the other nuclei in the system. The two-center one-electron integrals are written as a product of the corresponding overlap integral multiplied by the average of two atomic "resonance" parameters, β. There are only five types of one-center two-electron integrals surviving the NDDO approximation within a sp-basis $G_{ss}, G_{sp}, H_{sp}, G_{pp}, G_{p2}$. The G-type parameters are Coulomb terms, while the H parameter is an exchange integral. The G_{p2} integral involves two different types of p-functions (p_x, p_y, p_z). The orbital exponents are represented as Slater type and given as ζ_s, ζ_p. The core-core repulsion terms are given through the parameters a, b, c and due the inherent approximation in the NDDO method and the defect thereof in PM6 additional diatomic parameters were introduced ($a_{HH}, x_{HH}, a_{OO}, x_{OO}, a_{OH}, x_{OH}$).
The default set of PM6 parameters for hydrogen is given in table 3.7 and the according set for the oxygen parameters of the PM6 methods in table 3.8. The diatomic repulsion

parameters are listed in table 3.9 with the fitted default values.

PM6 parameters for H	
U_{ss}	-11.247
β_s	-8.3530
ζ_s	1.2686
g_{ss}	14.449
a	0.024184
b	3.0560
c	1.7860

Table 3.7 – Standard PM6 values for hydrogen [33].

PM6 parameters for O	
U_{ss}	-91.679
U_{pp}	-70.261
β_s	-65.653
β_p	-21.623
ζ_s	5.4218
ζ_p	2.2710
g_{ss}	11.304
g_{sp}	15.807
g_{pp}	13.618
g_{p2}	10.333
h_{sp}	5.0108
a	-0.017771
b	3.058310
c	1.8964

Table 3.8 – Original parameters for PM6 oxygen

diatomic parameters for OH	
a_{HH}	3.5409
x_{XX}	2.2436
a_{OO}	2.6240
x_{OO}	0.53511
a_{OH}	1.2609
x_{OH}	0.19230

Table 3.9 – diatomic parameters for hydrogen and oxygen as originally fitted [154].

3.6.2 Results

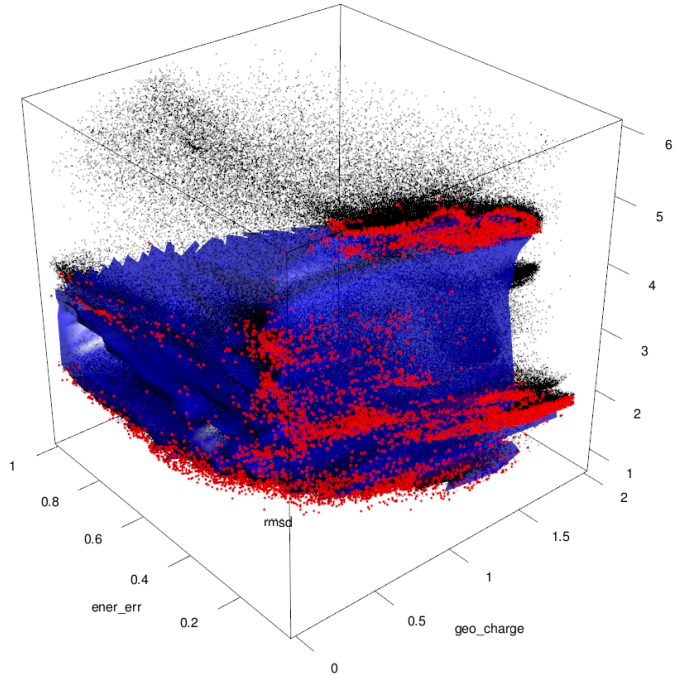


Figure 3.7 – Reduced objective space where water geometry error f_3 and charge error f_4 are mapped to one dimension. In red are represented the parameters that belong to the Pareto set. The blue surface is the interpolated Pareto front and the evaluated vectors are shown as black dots.

For an initial visual inspection of the results of the optimization run, we plot the reduced objective space in $\mathbb{R}^3$. Therefore the water monomer geometry error (f_3) and charge error (f_4) are mapped as described earlier. The resulting plot is shown in Figure 3.7 where the blue surface represents the Pareto front and the small dots the individuals or solution vector x_k with resulting objectives. Coloured in red are the solutions belonging to the Pareto set and illustrated in Figure 3.7. Thus, it is difficult to draw any conclusion from the output and constitute a further selection. Hence, the scatter-plot matrix is used to narrow down the selection of the solution space. As a first perceptive we shall focus on the scatter diagram of total RMSD and enthalpy error as shown in Figure 3.8 where the green dots represent the members of the Pareto set and in red a interpolation is drawn for the 2-dimensional Pareto front of the two objectives. This plot suggests that a selection for the RMSD and enthalpy error can be made. Thus, the solution we are likely to be interested in is in the area below RMSD < 2 and enthalpy error < 0.3. This is because only marginal improvements in either objective can be made outside this region. For decision making purposes we have used both, the total RMSD and the total enthalpy error which must be in mentioned boundary. The ability to set these additional determinations is one of the advantages of the multi-objective optimization process as one can not know in advance how the objectives will appear in the optimization process itself.

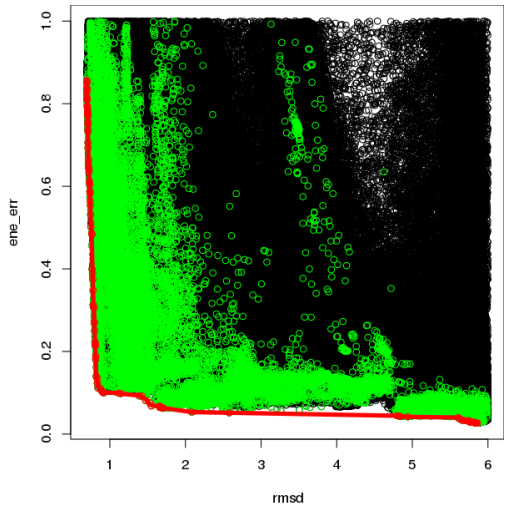
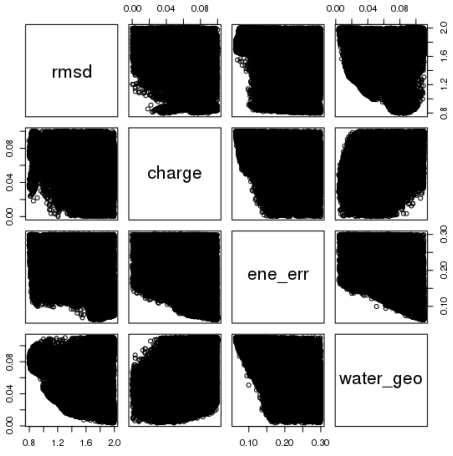
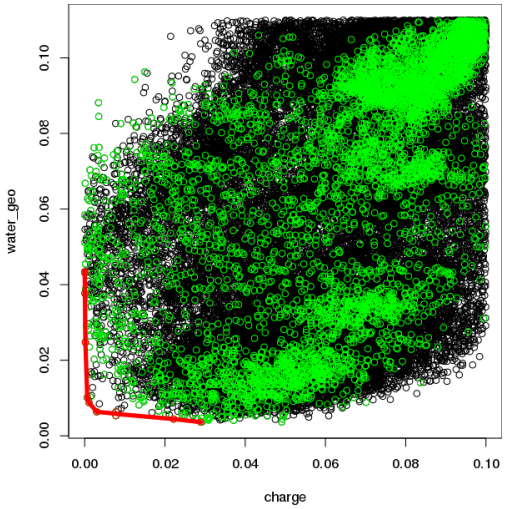


Figure 3.8 – The total RMSD error [Å] is plotted versus the total relative enthalpy error [kcal/mol]. The green dots represent the projected Pareto set and the red line is the approximated 2D Pareto front.

Thus, by applying proposed selection to the output data we again plot the scatter matrix as shown in Figure 3.9a where all evaluated data sets are plotted, including the Pareto set. This time selecting the relation of total RMSD and water monomer geometry for



(a) Scatter-plot matrix for all objectives



(b) Charge error versus water monomer geometry

Figure 3.9 – Solution vectors $x_k$after initial selection

further analysis. In plot **??** the objective RMSD is plotted against the error in geometry of the water monomer where the black dots are all solutions from the optimization output, green dots represent the Pareto set and the red line the fitted 2-dimensional Pareto front for the objectives f_2, f_3. The eventual option would be to cut all water geometries that exceed an error of 0.02 [Å]. Thus, the final selection boundaries for the PM6 parameter

sets for the 4 objectives are:

- f_1 ene_err < 0.3 kcal/mol
- f_2 RMSD < 2 Å
- f_3 water_geo < 0.01 Å
- f_4 charge < 0.02 e

As described earlier, from the resulting set we performed a k-median clustering and decided to employ $k = 4$ such that a small, representative sample set is obtained from the large Pareto front.

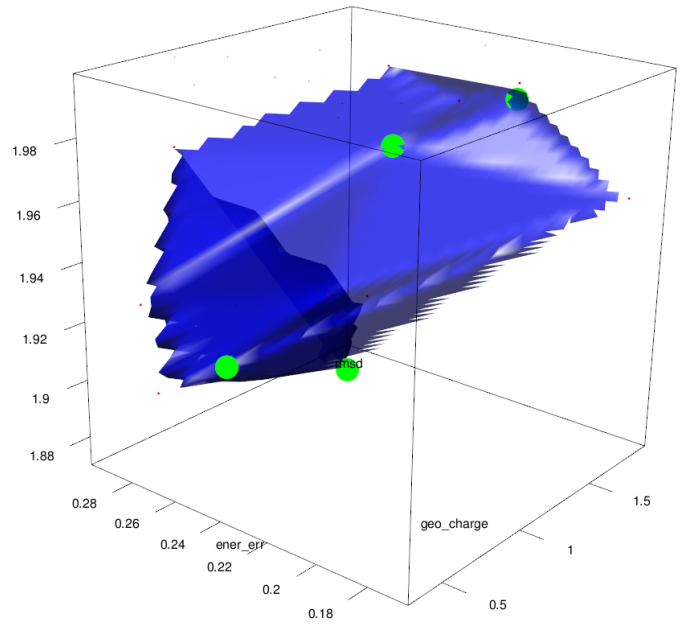


Figure 3.10 – Resulting parameter vectors x_k with according objectives after the selection. The water geometry was combined with the charge error to render a 3D plot for visual inspection. Red dots represent the Pareto set and the blue surface the interpolated Pareto front whereas the black dots are evaluated parameters. In green the selected parameters from the k-median clustering are shown.

Table 3.10 summarizes the resulting candidates, where all the 27 parameters of the PM6 method are shown as well as the 4 objectives, respectively the total enthalpy error (ene_err) [kcal/mol], total geometry error (RMSD) [Å], monomer geometry error (water_geo) [Å] and charge error (charge) [e] for the QCE reference data set alongside with the detailed enthalpy errors and geometry errors for the 8 structures in the set. For the 4 sets obtained from the k-median clustering, a short bulk water simulation in the canonical ensemble was performed. After a short equilibration phase ≈ 10 [ps], the oxygen-oxygen radial distribution function (rdf) was calculated and compared with the

	1	2	3	4
Uh_ss	-11.1488	-11.1647	-11.1949	-11.1543
betah_s	-8.1790	-8.1866	-8.1886	-8.1828
Zh_s	1.3278	1.3283	1.3286	1.3265
gh_ss	14.0171	14.0273	14.0369	14.0097
ah	0.0244	0.0238	0.0239	0.0238
bh	3.2790	3.2554	3.2541	3.2913
ch	1.7685	1.7916	1.8014	1.8014
Uo_ss	-92.6815	-94.0108	-93.3139	-93.9762
Uo_pp	-73.3399	-73.9947	-73.6156	-73.6218
betao_s	-65.6594	-66.0067	-65.9796	-66.1172
betao_p	-23.4438	-23.4025	-23.3615	-23.2638
Zo_s	4.9249	4.9095	4.9333	4.9036
Zo_p	2.0439	2.0442	2.0439	2.0447
go_ss	11.6973	11.8355	11.7903	11.8728
go_sp	14.7819	14.8956	14.9184	14.8799
go_pp	12.7095	12.6064	12.6603	12.6188
go_p2	11.1047	11.1840	11.1015	11.1239
ho_sp	5.4317	5.4100	5.4373	5.3895
ao	-0.0180	-0.0178	-0.0177	-0.0178
bo	3.2590	3.2779	3.2682	3.2697
co	1.7101	1.7134	1.7239	1.7086
x_hh	2.0551	2.0623	2.0561	2.0573
x_oh	0.2016	0.2004	0.1998	0.2000
x_oo	0.5163	0.5198	0.5199	0.5202
a_hh	3.5152	3.4757	3.4704	3.4821
a_oh	1.2678	1.2479	1.2575	1.2631
a_oo	2.4494	2.4757	2.4861	2.4984
ene_err	0.2577	0.2061	0.2727	0.2361
rmsd	1.9091	1.9881	1.8778	1.9728
water_geo	0.0063	0.0084	0.0098	0.0082
charge	0.0078	0.0189	0.0120	0.0135
ene_0	0.0000	0.0000	0.0000	0.0000
ene_1	0.0373	0.0442	0.0257	0.0341
ene_2	0.1212	0.1107	0.1287	0.1234
ene_3	0.0236	0.0122	0.0257	0.0176
ene_4	0.0217	0.0105	0.0231	0.0163
ene_5	0.0165	0.0099	0.0234	0.0148
ene_6	0.0162	0.0069	0.0198	0.0115
ene_7	0.0213	0.0117	0.0263	0.0185
rmsd_0	0.0063	0.0084	0.0098	0.0082
rmsd_1	0.1022	0.1010	0.1039	0.1033
rmsd_2	0.2558	0.2496	0.2486	0.2573
rmsd_3	0.3193	0.4525	0.4464	0.4544
rmsd_4	0.1706	0.1663	0.2218	0.1653
rmsd_5	0.2355	0.2285	0.2336	0.2346
rmsd_6	0.2625	0.2477	0.2465	0.2568
rmsd_7	0.5569	0.5343	0.3673	0.4930

Table 3.10 – The resulting candidates of the k-median clustering. After a short NVT simulation and evaluation of the radial distribution function, the last candidate was selected.

experimental data. From visual inspection of the oxygen-oxygen rdf the final selection of the parameter set was carried out. Thus, the resulting parameter set for the PM6 method are presented in Table 3.11 along with the original values and the change in percent. It is interesting to see that large deviation is encountered for the Slater atomic orbital exponent both, for oxygen Z_s^o,Z_p^o and hydrogen Z_s^h. Since hydrogen-hydrogen core-core repulsion is not frequently encountered in liquid water, the large change in x_{hh} is likely to be trying to correct the oxygen-hydrogen interaction although a_{oo} is also reduced. For the mono-atomic parameters a, b, c it can be assumed that especially through the change in the oxygen parameter, a narrower basin of attraction with a bigger radius is employed. However, it is difficult to asses individual parameters specifications and a rigorous sensitivity analysis of the PM6 parameters should be performed to fully understand and asses the underlying connections.

	original	final	change in %
Uh_ss	-11.2470	-11.1543	-0.8242
betah_s	-8.3530	-8.1828	-2.0376
Zh_s	1.2686	1.3265	4.5646
gh_ss	14.4487	14.0097	-3.0381
ah	0.0242	0.0238	-1.4701
bh	3.0560	3.2913	7.7018
ch	1.7860	1.8014	0.8643
Uo_ss	-91.6788	-93.9762	2.5059
Uo_pp	-70.4609	-73.6218	4.4859
betao_s	-65.6351	-66.1172	0.7344
betao_p	-21.6226	-23.2638	7.5902
Zo_s	5.4218	4.9036	-9.5570
Zo_p	2.2710	2.0447	-9.9631
go_ss	11.3040	11.8728	5.0316
go_sp	15.8074	14.8799	-5.8679
go_pp	13.6182	12.6188	-7.3386
go_p2	10.3328	11.1239	7.6565
ho_sp	5.0108	5.3895	7.5571
ao	-0.0178	-0.0178	0.0432
bo	3.0583	3.2697	6.9112
co	1.8964	1.7086	-9.9035
x_hh	2.2436	2.0573	-8.3015
x_oh	0.1923	0.2000	3.9962
x_oo	0.5351	0.5202	-2.7794
a_hh	3.5409	3.4821	-1.6609
a_oh	1.2609	1.2631	0.1744
a_oo	2.6240	2.4984	-4.7877
ene_err	0.9130	0.2361	-74.1426
rmsd	5.8544	1.9728	-66.3017
water_geo	0.0313	0.0082	-73.9577
charge	0.0416	0.0135	-67.4659
ene_0	0.0000	0.0000	
ene_1	0.1511	0.0341	-77.4133
ene_2	0.1418	0.1234	-13.0077
ene_3	0.1220	0.0176	-85.6036
ene_4	0.1194	0.0163	-86.3422
ene_5	0.1359	0.0148	-89.0810
ene_6	0.1242	0.0115	-90.7807
ene_7	0.1188	0.0185	-84.4557
rmsd_0	0.0313	0.0082	-73.9577
rmsd_1	0.0747	0.1033	38.3160
rmsd_2	0.6150	0.2573	-58.1647
rmsd_3	1.0500	0.4544	-56.7247
rmsd_4	1.1868	0.1653	-86.0721
rmsd_5	0.4057	0.2346	-42.1709
rmsd_6	1.1461	0.2568	-77.5900
rmsd_7	1.3449	0.4930	-63.3447

Table 3.11 – Comparison of the final parameters with the original PM6 values and the change in percent

3.7 Parametrizing DFTB3 γ-function

3.7.1 γ-function and Parameters

As described in section 2, γ_{ab} is derived from the assumption that the electron-electron interaction in the second-order terms of the total DFTB3 energy can be evaluated from the interaction of two exponentially decaying charge densities, in which the exponent τ_a is a measure for the extension of the atomic charge density. Further, the on-site interaction γ_{aa} corresponds to the electron self-interaction on the atom and hence can be expressed via the Hubbard parameters U_a, which are twice the chemical hardness ν_a:

$$\gamma_{aa} = U_a = 2\nu_a \tag{3.7}$$

This leads to the assumption that there is an inverse correspondence between the size of an atom, $1/\tau_a$, and its chemical hardness parameter, U_a [146, 120]. In a recent work of Politzer and co-workers [171], various sets of covalent radii were compared with respective chemical hardness value and an reasonable agreement has been found. However, large deviation has been found in particular for the hydrogen atom.
Because γ_{ab} approaches the value $\gamma_{aa} = U_a$ at short distances, the poor relation between its size and the chemical hardness for H means that modifications have to be made for γ_{ab} for all X-H pairs. In principle, this could be done by modifying the value of U_H for hydrogen according to its atomic size, which would, however, make the on-site interaction on H, γ_{HH}, inconsistent with its chemical hardness.
In the standard implementation of the DFTB3 method, γ_{ab} has the form

$$\gamma_{ab} = \frac{1}{R_{ab}} - S \tag{3.8}$$

with S being a short-range function that leads to the desired limit for small inter-atomic distances. Since the hydrogen atom size according to $r_c = 6/(16U_H)$ is too large, the density overlap is overestimated and therefore the electronic interaction starts to deviate from $1/R_{ab}$ too early. To correct for this, Yang et al [120] proposed an additional damping term for the X-H pairs as

$$\gamma_{aH} = \frac{1}{R_{aH}} - S\ exp\left[-\left(\frac{U_a + U_H}{2}\right)^{\zeta} R_{aH}^2\right] \tag{3.9}$$

Where the single parameter ζ in the exponent is fitted to appropriate reference systems [146]. This modified γ_{aH} function has a significant impact on hydrogen bonding and hence Yang et al [146] systematically optimized the parameters (including the Hubbard derivatives U_a^d for the third order approximation) by fitting based on the binding energies and proton affinities of a set of gas-phase compounds that are of general biological interest.
To this correction the DFTB3 binding energy for the water dimer could be reduced to

−5.4 kcal/mol which is very close to the expected value of −5.0 kcal/mol [172]. Hence, this adoption was widely incorporated [173, 174, 175] but ignoring the crucial fact that this correction only holds for *binding energies*. Therefore, plotting all the objectives (enthalpy [kcalmol] and RMSD [Å] error for the QCE set and geometry error [Å] for the water monomer as well as Mulliken charge error [e] for the water monomer) against different values of ζ reveals an intrinsic problem of the proposed damping term as shown in Figure 3.11.

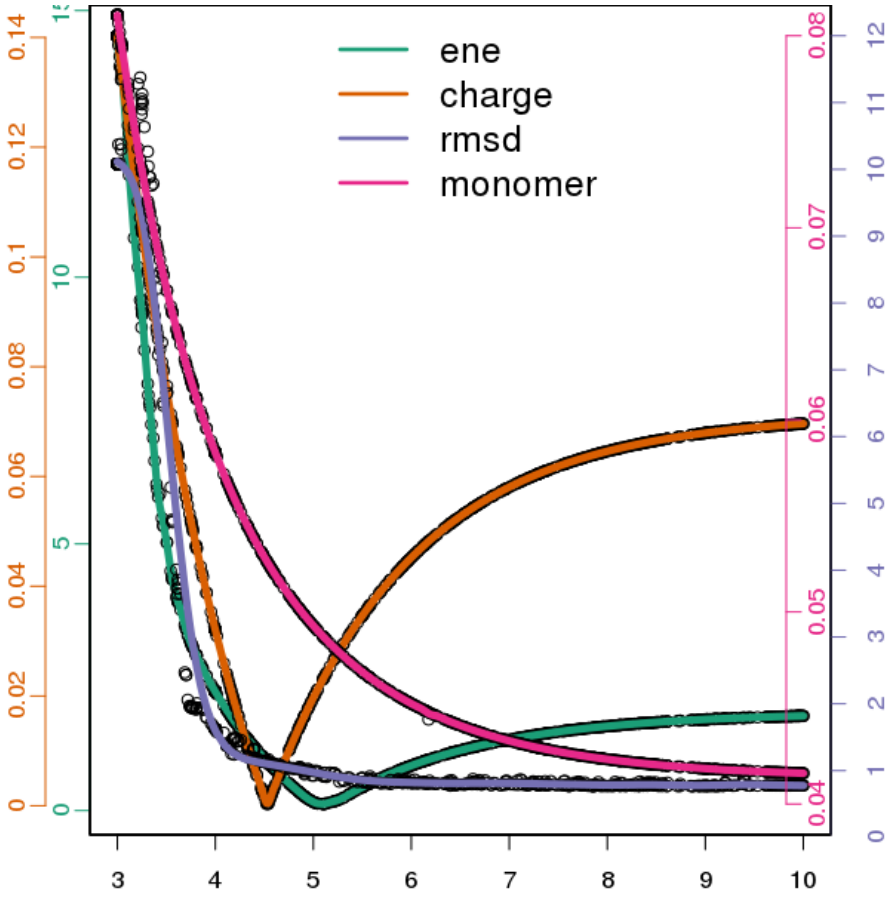


Figure 3.11 – A overview of the effect of the γ-function by Yang [146] with different values for ζ. The error in enthalpies are given in [kcal/mol], RMSD [Å], monomer geometry error [Å] and charge error [e]. Although the energies can be improved compared to the standard calculations (limes) this comes at cost of inferior geometries.

Although there is a clear minimum for the enthalpy error around $\zeta \approx 5$ one can see that this only comes at a cost of larger deviation of the geometries. This suggests that obtaining better binding energies at the same time worsens the forces and hence puts the damping function proposed by Yang et al at question. That this can lead to drastic consequences as was first published by Choi et al [175] without attributing the large voids that formed during the simulation to the damping function of Yang [146]. Figure 3.12 shows the snapshots of the bulk water, hydrated hydroxide, and hydrated excess proton system after

50 picoseconds of NVT molecular dynamics simulations at a density of 0.97 g/mL. As stated in [175], this phenomenon observed occurs primarily due to the over-coordinated water molecules which in turn is an artefact of the damping function. Therefore, a closer

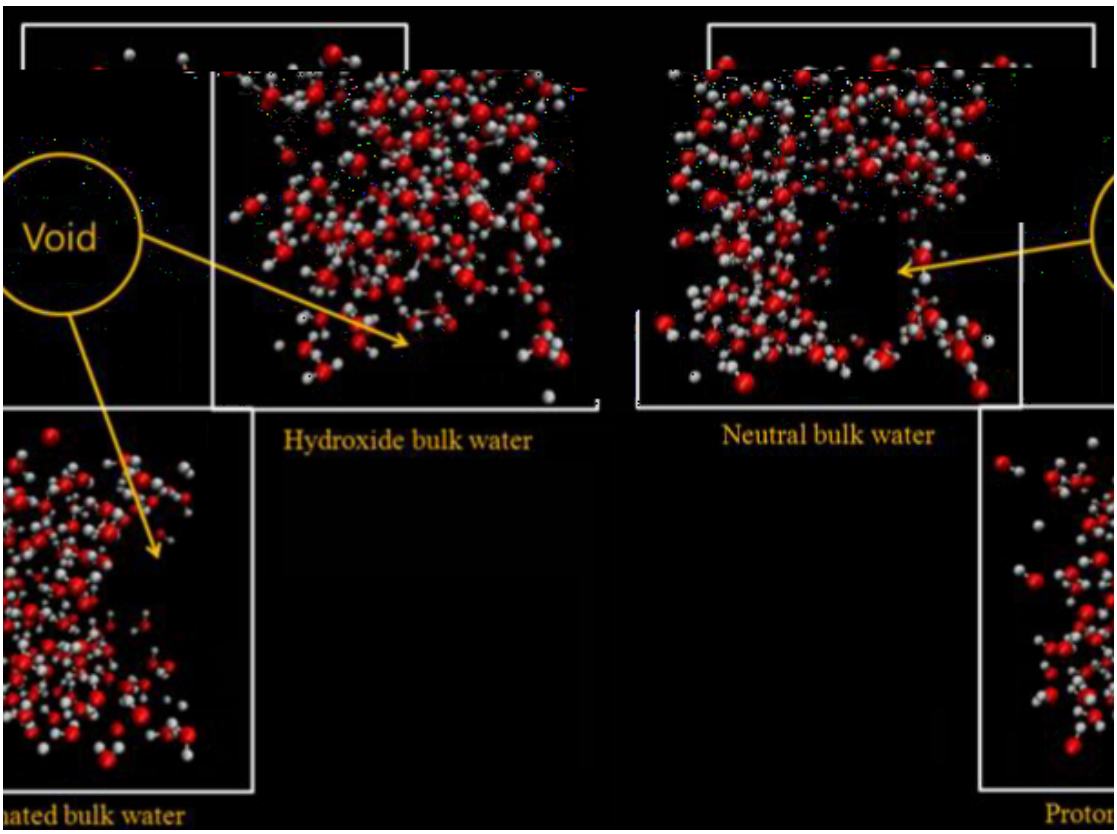


Figure 3.12 – Snapshots of the neutral, hydroxide, and protonated bulk system after 50 ps of the MD time evolution using the DFTB method (with γ_{aH} correction). The voids are detected in all simulations [175].

look at the γ function was taken and a more flexible damping correction of the form

$$cor = -a \cdot exp(\frac{-(R-b^2)}{c^2}) + 1 \tag{3.10}$$

proposed, where a, b, c are parameters to be fitted. The basic idea starts with the assumption that the short and long term interaction should not be altered heavily. Hence, a Gaussian function is introduced with flexible width c, mean b and amplitude a. The resulting gamma function has the final from of

$$\gamma_{aH} = \frac{1}{R_{aH}} - S \left[1 - a \cdot exp(\frac{-(R-b^2)}{c^2})\right] \tag{3.11}$$

As comparison, the functional forms of the standard γ_{HO} interaction (black) is plotted against the corrected Yang (red) and the proposed (green) relations (Figure 3.13). Instead

of a switch function as suggested by Yang and plotted with solid blue line, a Gaussian function is proposed (dashed blue line) to allow better interaction energies and forces as shown in Figure 3.13.
Hence, the resulting task is to optimize the parameters a, b, c such that good agreements in the desired objectives can be made. The set-up described earlier is used to perform this multi-objective optimization problem and results are presented in the following subsection.

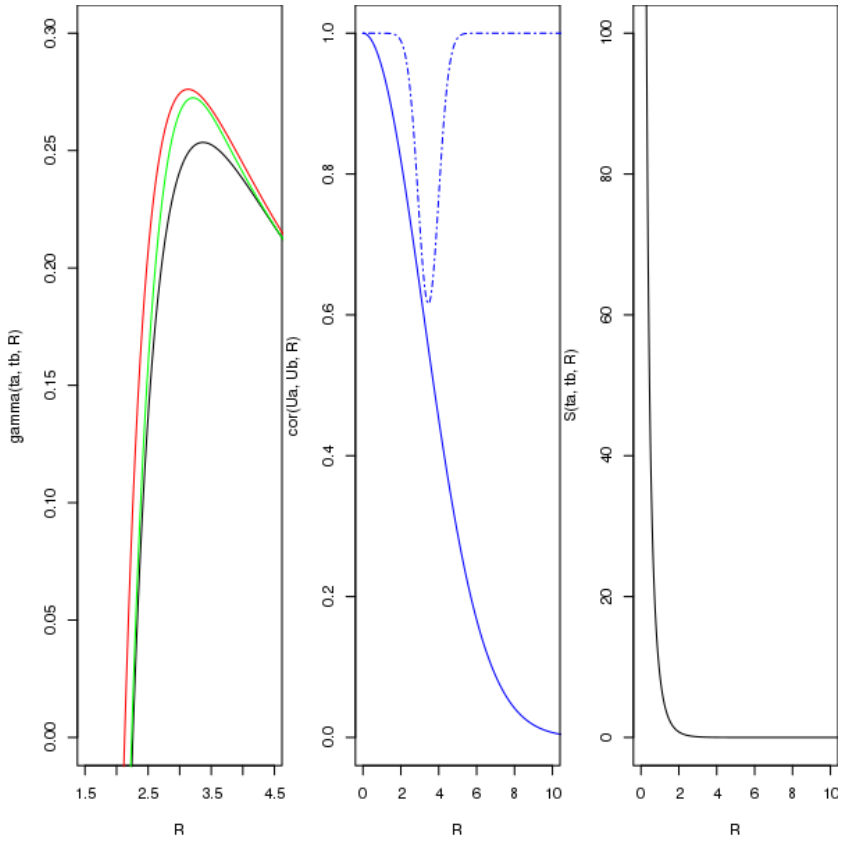


Figure 3.13 – Different γ-functions are plotted on the left, where the black line represents the original, the red line the function proposed by Yang [146] and in green our suggestion. The actual correction functions are shown in the middle where the blue line is the switch function by Yang and the blue dotted line our proposed correction. On the right the short range function S is plotted.

3.7.2 Results

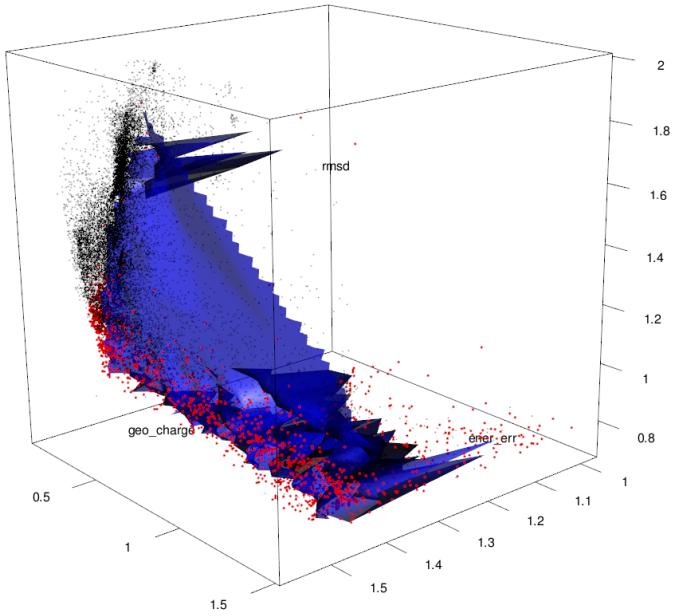


Figure 3.14 – Reduced objective space where water geometry error f_3 and charge error f_4 are mapped to one dimension. The red dots represent the parameters that belong to the Pareto-set. The blue surface is the interpolated Pareto-front and the evaluated vectors are shown as black dots.

The results from the multi-objective optimization are plotted as reduced objectives as shown in Figure 3.14 where again the blue surface represents the Pareto-front and the small dots the solution vector $x_k = \{a_k, b_k, c_k\}$ with corresponding objectives. Coloured in red are the solutions of the Pareto-optimal-set and to gather more information the scatter-plot matrix is investigated in Figure 3.15. There are only marginal changes in charge errors and water monomer geometry as interfered from the scatter-plot matrix. This is somewhat not surprising since the γ-function correction is limited. Thus especially to the properties such as monomer geometry and partial atomic charge, small changes can be expected. The focus is shifted to the total geometric and energetic error since those properties are more affected by the γ-function through a stronger hydrogen bond behaviour. From the closer inspection of the total RMSD versus enthalpy error plot in Figure 3.16 it becomes clear to apply a selection of RMSD < 0.9 Å to the actual data. This plot also indicates, that the γ-correction can not further improve on total energy error beyond about 0.2 kcal/mol. A review of the $k = 4$ medians of the clustering algorithm revealed some correspondence of over-structured first peak hight in the radial distribution function of the oxygen-oxygen in the initial NVT simulation. Hence, applying a stricter selection of RMSD < 0.75 Å was enforced. As mentioned above, the charge and monomer geometry error are marginally altered and hence from the trimmed set we sorted according to minimal energy error and the final solution vector was found to be $x_{final} = \{0.7700578,\ 0.3843653,\ 3.452199\}$.

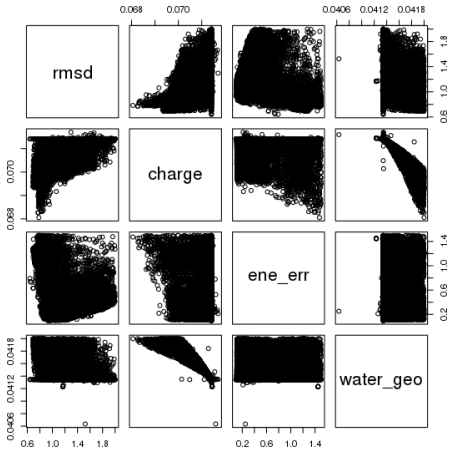


Figure 3.15 – Scatter-plot matrix of all objectives of the optimization run for the γ function fitting

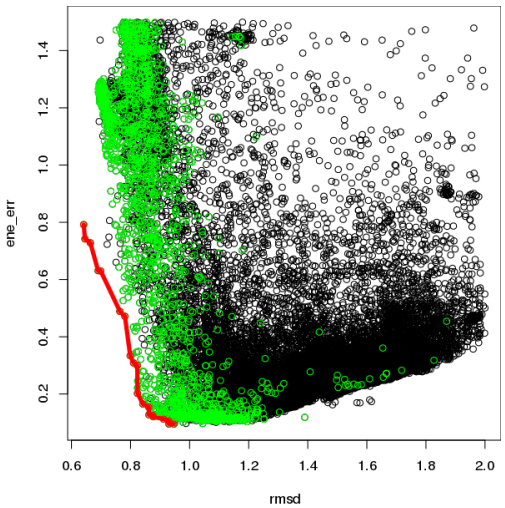


Figure 3.16 – The total RMSD error versus the total enthalpy error for the entire set of solutions to the γ optimization problem

3.8 Bulk water properties

A major challenge in condensed matter simulations is understanding the structural and electronic properties of liquid water at ambient conditions. Water is a crucial ingredient for a large variety of systems of importance, from basic chemistry, biology, and physics, as well as in the applied fields of catalysis and energy production. The water molecule is somewhat special in respect that it possess a large dipole moment and polarizability, is a multiple hydrogen donor and acceptor and can easily build network structures. The total cohesive energy in the condensed phase is therefore a sum of many weak interactions and the theoretical model face the challenge to describe many different effects. The development of empirical potentials for water [89, 176] improved the understanding of water's behaviour and properties [177, 178, 179]. However, empirical models lack transferability and might fail if used under conditions outside of their fitting range and moreover, when water takes an active role in a chemical process, the electronic properties of the water molecule need to be taken into account. In this respect, first-principle methods offer the only solution.The capability to reproduce properties of complex systems such as liquid water can therefore be used to judge the sophistication and predictive power of a given model. Many DFT based simulation of bulk water have been reported in the literature, and in this context three main methods of sampling the phase space can be recognized [180]; Car-Parrinello molecular dynamics (CPMD) and derivatives thereof [181, 182, 183, 184, 185], Born-Oppenheimer molecular dynamics (BOMD) [186, 187, 184] and Monte Carlo (MC) sampling [188].
Between these extreme approaches of classical force fields and ab-inito description of water are the semi-empirical quantum chemical methods [189]. Since they are still quantum-mechanical at their core, these methods contain the physics of polarizability that is difficult to capture otherwise. At the same time, these methods are also faster than traditional electronic structure approaches.
However, to date only few reviews addressing the performance of semi-empirical methods reproducing bulk water at ambient conditions [173, 158] exists. This is somewhat surprising especially since DFTB is predominantly used in biochemical simulations [190, 85, 191] where it is crucial to be able to reproduce ambient water conditions for enzymes or small proteins. Hence, in this section we compare certain properties of bulk water at ambient conditions of the standard PM6 and DFTB3 with the optimized methods introduced in this manuscript. For that purpose, density, radial distribution function, surface tension and diffusion coefficients are compared.

3.8.1 Computational Details

All calculations presented have been performed with the CP2K program [149]. The classical force-fields TIP3P [192] and SPC/E [193] are employed with the standard reference parameters. The system sizes for the classical force-fields correspond to 256 water molecules whereas the model system for the semi-empirical consist of 115 water molecules

in a cubic simulation cell under periodic boundary conditions (PBC). Initial configuration for the semi-empirical bulk simulations were taken from extensively equilibrated TIP3P runs at ambient condition. The NVT ensemble simulation were carried out at a density of 999.5 kg/m^3 corresponding to a square cell 15.1 Å on a side. Simulations were run at 300 K using the Nose-Hoover thermostat [44,45] and an integration time step of 1 fs was applied. Further NPT simulations were carried out at the thermodynamic constraint set to ambient conditions, that is, T = 300 K and p = 1 bar. Additional NVE simulations were carried out at respective equilibrium densities obtained from the NPT simulations. The atomic cut-off radius for the evaluation of the Coulomb integrals was set to 12 Å and a modified Klopman-Dewar-Sabelli-Ohno screening (KDSO-D) [158] was employed. The same cut-off value was used for the exchange integrals. The periodicity, and hence the electrostatic long-range part was handled through the usage of multi-pole Ewald summation schemes. The maximum level of multi-poles expansion used for the electrostatics was set to use up to the Quadrupole term. For the DFTB3 calculations smooth particle mesh methods using beta-Euler splines were employed. The usual trajectory length was targeted to be 100 ps with 5 ps equilibration. The computational details for the Monte Carlo simulations of the ab initio reference calculations can be found [194].
The NVT slab calculations needed to model the ambient liquid-vapour interface and hence were performed in a similar fashion to the bulk water NVT calculations (similar set-up to Murdachaew et al [158]) with the exceptions that 512 water molecules were used and the dimension of the rectangular box were 19.7 Å(x) × 19.7 Å(y) × 100 Å(z). This is larger than proposed by Murdachaew but necessary for the fluctuation surface analysis. The simulation time for the big system was targeted for 300 ps for both semi-empirical simulation and 6 ns for the force field methods.

3.8.2 Radial Distribution Functions

The radial distribution functions, g_{OO}, g_{OH}, and g_{HH} are commonly used when the structure of the liquid water is studied [195]. These intermolecular partial pair correlation functions for liquid water at 25° C were determined from neutron diffraction data by Soper et al [196, 197].The old [196] and new [197] results for g_{OO}, g_{OH}, and g_{HH} are in good agreement, except that the first O-H peak at 1.8 Å is increased by about 14% compared the previous analysis. The differences probably represent the currently available accuracy in determining the site-site pair correlation functions for water. The radial distribution functions, g_{OO}, g_{OH}, and g_{HH} are easy to calculate from molecular dynamics data and are generally used when different water models are compared with experimental.

3.8.2.1 Results from PM6 and DFTB3 simulations

The radial distribution function computed from our simulations for the PM6, PM6$_{optim}$, DFTB3 and DFTB3$_{optim}$ are compared with experimental data and ab-inito results. For that, BLYP (from the name Becke for the exchange part and Lee, Yang and Parr for the correlation part) density functional theory (with dissperison correction D3) calculations were employed. Details of the Monte Carlo simulation for the ab-inito calculations can be found [170] and [198]. For a comparison on the PM6, in a recent publication [199] Welborn re-optimized PM6 with a force-matching approach and hence the resulting parameters are used to perform the same canonical ensemble simulation. The entire analysis is carried out with the help of Travis [200], an analyser and visualizer for Monte Carlo and Molecular Dynamics Trajectories. In Figure 3.17 we show the results of the oxygen-oxygen radial distribution function g$_{oo}$ for the PM6 methods with different parameter sets (original, Welborn [199] and our optimized set) as well as the original DFTB3 and the γ-corrected version with parameters a, b, c obtained from our optimization runs. It becomes clear that the original PM6 method has features shifted dramatically outwards and thus fails in predicting the correct structure. Also, the first peak is slightly shifted to lower distances wrongly packing the bulk water. The parameter set from Welborn et al [199] show slight over-structuring in the oxygen-oxygen radial distribution function but are in much better agreement with the experimental results as the original parameter set. Although stating in their publication, Welborn et al [199] claimed that force-matching is necessary to properly capture the basic structure of water, we could show that the small QCE reference data set is enough to reproduce a reasonably close approximation to the experimental curve. Especially the mid and long range part agree quite well with the reference curve whether the first peak is slightly over-pronounced and is generally attributed to the lack of quantum effects in the simulations [170].
For the original DFTB3 method similar offsets as for the original PM6 methods are obtained. The g$_{oo}$ is shifted to the right and the mid and long range parts are almost flat and unstructured. The γ-correction improves drastically the mid and long range parts at the coast of a well over-pronounced first peak attributed to a more packed oxygen-oxygen bulk behaviour.
Both BYLP Monte Carlo simulation results are shown in Figure 3.17. Although the sampling was carried out in NPT one can deduce that the standard BLYP without dispersion correction performs rather poorly and is highly over-structured but the addition of the missing theory reduces this error.
Assessing only the oxygen-oxygen radial distribution function, we could show that the optimized parameters for PM6 show even better agreement to experimental results than elaborate and time-consuming ab-inito methods such as BLYP and BLYP-D3.

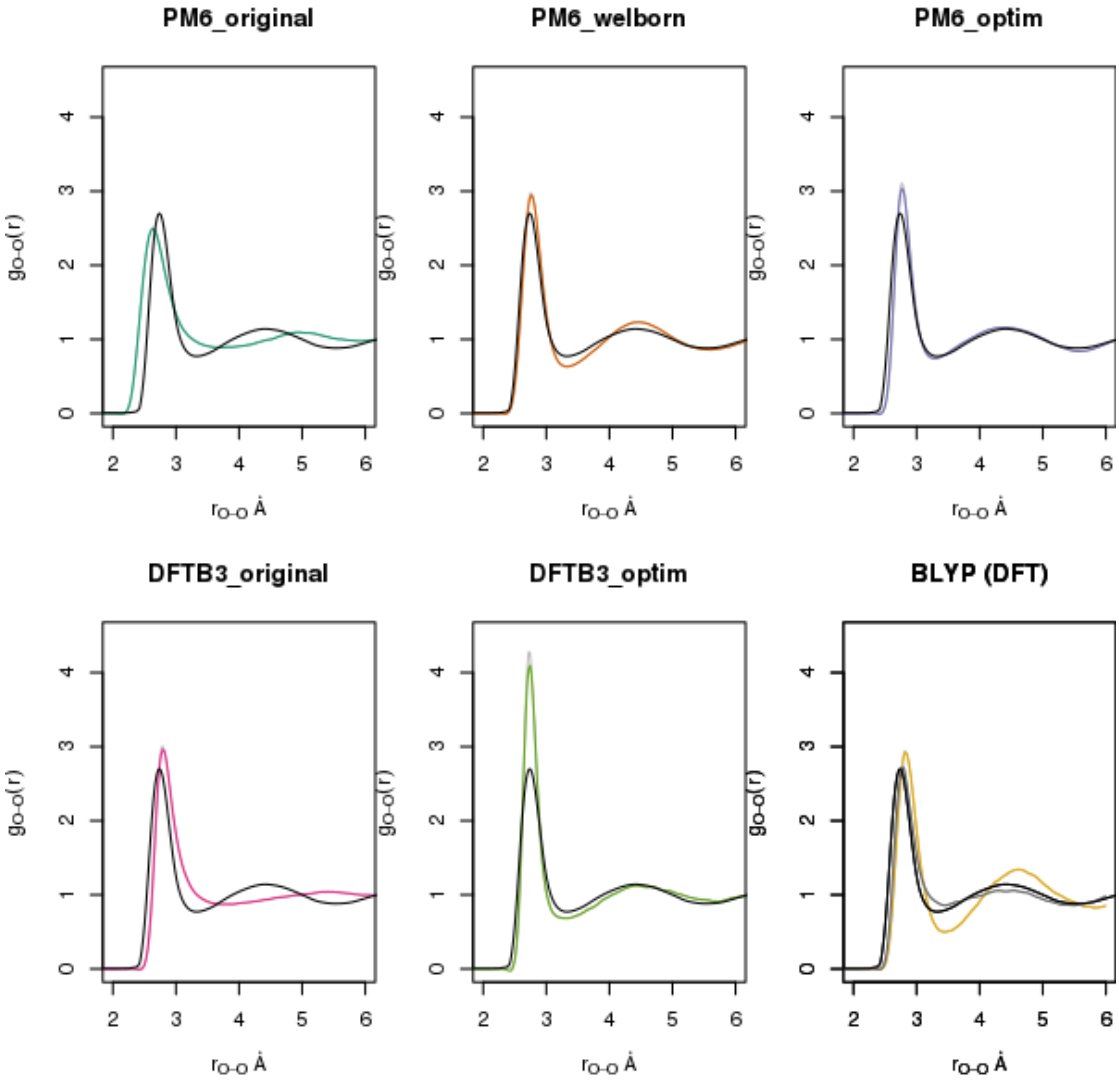


Figure 3.17 – Different oxygen-oxygen radial distribution functions from NVT simulations at reference density $\sigma = 999.5$ kg/m^3 and T = 300 K. The BLYP results of the NPT Monte Carlo simulations are shown additionally for comparison purpose [170, 198]

3.8.3 Density

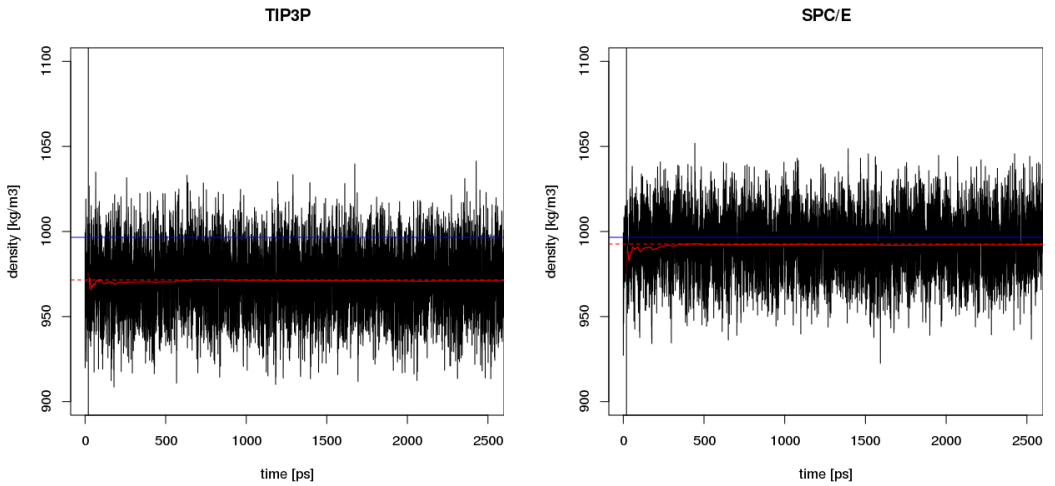
The density of water is approximately one gram per cubic centimetre. It is dependent on its temperature, but the relation is not linear and uni-modal rather than monotonic (see table 3.12).

Temp (° C)	+100	+80	+60	+40	+30	+20	+15	+10	+4	0
Density (kg/m^3)	958.4	971.8	983.2	992.2	995.65	998.21	999.1	999.7	999.97	999.84

Table 3.12 – water density for different temperatures

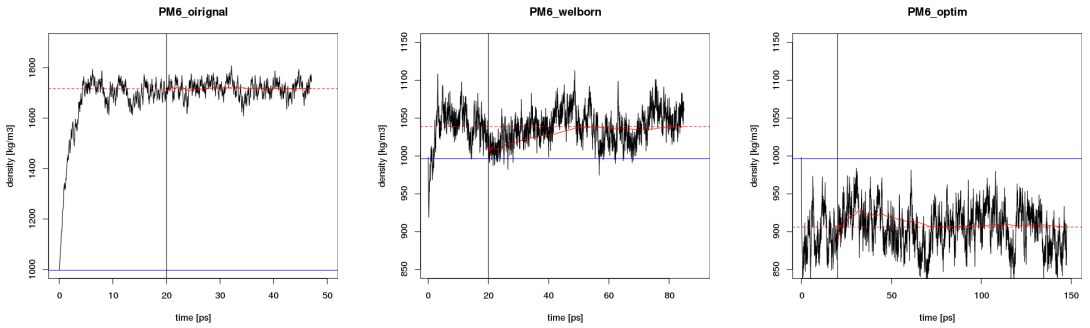
When cooled from room temperature liquid water becomes increasingly dense, as with other substances, but at approximately 4 C°, pure water reaches its maximum density. As it is cooled further, it expands to become less dense. This unusual negative thermal expansion is attributed to strong, orientation-dependent, intermolecular interactions. Hence in the next subsection, the density is calculated for chosen semi-empirical methods at ambient condition. Thus, the pressure of 1 atm and Temperature of 300 K are maintained with standard Nosé-Hoover thermostat and barostat, see CP2K reference [149] and computational details for further information.

3.8.3.1 Results from PM6 and DFTB simulations



(a) TIP3P density simulation (b) SPC/E density simulation

Figure 3.18 – Force-field simulation at ambient conditions, T =300 K, p = 1 atm and resulting instantaneous density fluctuation and mean.

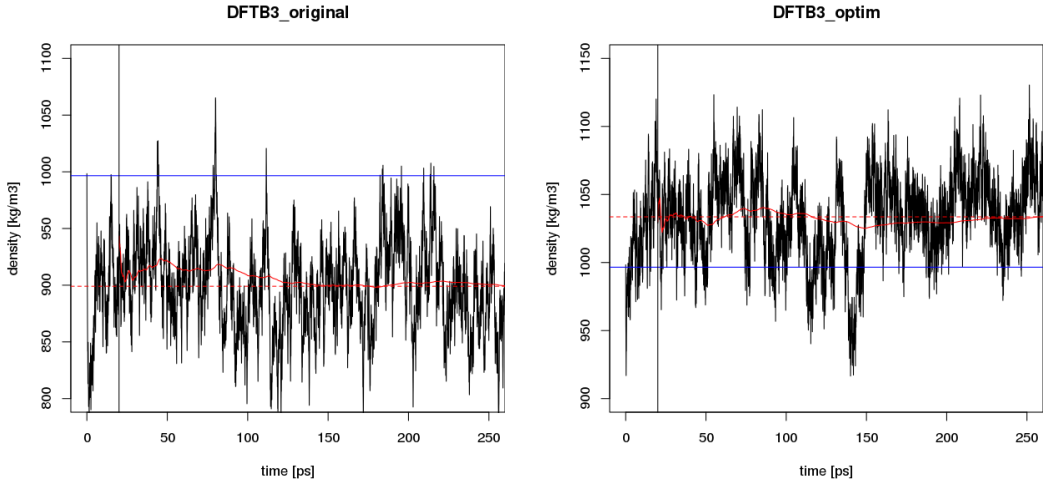


(a) Original PM6 simulation

(b) PM6 simulation with parameters from Welborn et al [199]

(c) PM6 with optimized parameter simulation

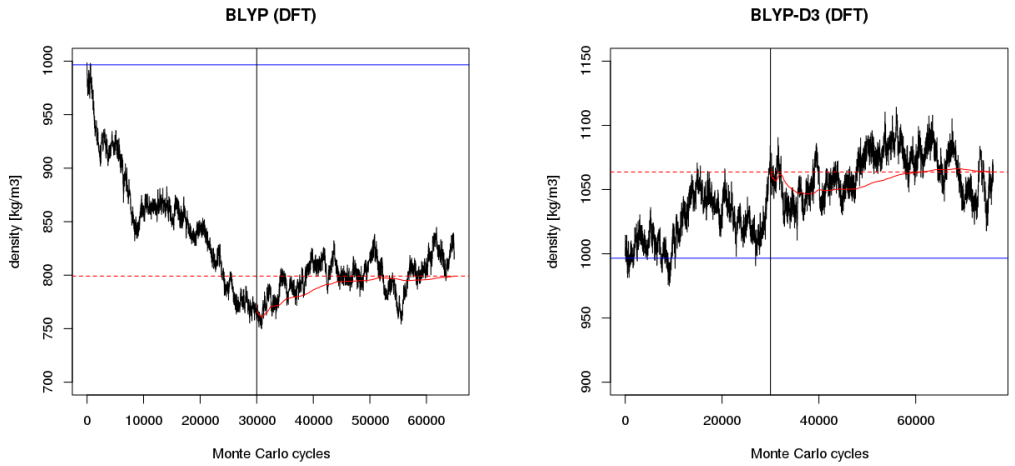
Figure 3.19 – Different PM6 semi-empirical simulations at ambient conditions, T = 300 K, p = 1 atm and resulting instantaneous density fluctuation and mean.



(a) DFTB3 original density simulation

(b) DFTB3 with proposed γ-function correction density simulation

Figure 3.20 – DFTB3 semi-empirical simulation at ambient conditions, T = 300 K, p = 1 atm and resulting instantaneous density fluctuation and mean.



(a) BLYP density Monte Carlo simulation (b) BLYP-D3 density Monte Carlo simulation

Figure 3.21 – Fluctuation of the instantaneous density as a function of the Monte Carlo cycles for the NPT simulation.

The value of the density of liquid water is obtained from averaging the fluctuating instantaneous density from the molecular dynamics simulation and Monte Carlo sampling for the ab-inito methods [170] starting after an equilibrium phase (20 ps or 30000 Monte Carlo cycles). The instantaneous density and the corresponding average are depicted in Figure 3.18a to 3.21b, while the calculated average value with the associated standard deviation are reported in Table 3.13. Our PM6$_{optim}$ results (shown in Figure 3.19c) for the density of liquid water at ambient condition are $\approx$ **905.95** kg/m^3 which is in moderate agreement with the experimental value of **995.65** kg/m^3 but much better than ab-inito BLYP values of **799.12** kg/m^3 [170] and comparable to dispersion corrected methods such as BYLP-D3 **1063.68** kg/m^3. For the DFTB3$_{optim}$ density (shown in Figure 3.21a and Figure 3.21b) **1033.57** an improvement over the standard DFTB3 density of liquid water of **899.06** can be identified. A possible reason for this is that the water density depends crucially on the medium to long range part of the potential [170]. Thus, the DFTB3$_{optim}$ density result is almost comparable to expensive and time-consuming ab-inito methods such as PBE0-ADMM-D3 [170] and is clearly in the range ob BLYP-D3 (shown in Figure 3.21b).

3.8.4 Surface tension

The ability of a model to predict interfecial properties, as is the case of the surface tension of the vapour-liquid interface can give indication about the quality and performance. The determination of the surface tension of water by computer simulation has been the subject of several studies [201, 202, 203, 204, 205, 206, 207, 208, 209, 210]. Usually, a slab of liquid is placed in contact with vapour and the surface tension is computed

	density [kg/m3]	standard deviation	relative error in %
TIP3P	971.5336	17.3834	-2.5110
SPC/E	992.6380	15.7346	-0.3933
PM6_original	1717.5825	30.7255	72.3517
PM6_welborn	1038.7378	21.5797	4.2327
PM6_optim	905.9513	26.7556	-9.0919
DFTB3_original	899.0629	39.7607	-9.7831
DFTB3_optim	1033.5710	30.9530	3.7142
BLYP	799.1156	17.3238	-19.8124
BLYP-D3	1063.6782	18.1213	6.7353
Experiment	996.5570		

Table 3.13 – Different density values obtained from NPT simulations compared to results from ab initio and experiment at ambient condition (T = 300K, p = 1 bar).

from a mechanical route which requires the calculation of the pressure tensor [211]. Unfortunately, the values of the surface tension from different authors differ in some cases considerably. Amongst other reasons for this discrepancies is that reliable values of the surface tension are only obtained after considering sufficiently large systems and long simulation runs [212].
In this subsection we consider simulations of the -liquid interface of $PM6_{optim}$, $DFTB3_{optim}$, classical forcefields such as TIP3P and SPC/E and determine the surface tension from the mechanical route as well as with a new method based on the fluctuation of the surface. The basic idea to use the surface fluctuation to determine the surface tension is described by Chandler [213] and briefly discussed below.

3.8.4.1 Mechanical Route to Surface Tension

A very popular and straightforward way to calculate the surface tension γ is to use the pressure tensor. For a planer interface perpendicular to the z axis, γ is given by

$$\gamma = \int_{\text{inf}}^{\text{inf}} dz \left[p_N(z) - p_T(z) \right] = L_z \left[\overline{p}_N - \overline{p}_T \right], \tag{3.12}$$

where $p_N(z)$ and $p_T(z)$ are the normal and tangential (local) components of the pressure tensor at position z, respectively. For a planar interface, p_N ideally does not depend on z and is equal to the vapour pressure. $\overline{p}_N$ and $\overline{p}_T$ in Equation 3.12 are macroscopic components of the pressure tensor defined in terms of the volume average of their local components counterparts [214]. Considering that the set-up of the simulation stabilizes two vapour-liquid interfaces, the working expression for the computation of the surface tension takes the simple expression

$$\gamma = \frac{L_z}{2} \left[\overline{p}_N - \overline{p}_T \right]. \tag{3.13}$$

Hence, a standard simulation in a canonical ensemble can be carried out where the pressure tensor is required to be calculated. Since this is a basic feature of the CP2K

package we refer to [149] for further details.

3.8.4.2 Surface Tension from Surface Fluctuation

The surface tension can also be estimated from the fluctuation of the instantaneous interface. Therefore, firstly a instantaneous liquid-vapour interface needs to be defined. The basic idea starts with the instantaneous density field at space-time point r, t

$$\rho(r,t) = \sum_i \delta(r - r_i(t)) \tag{3.14}$$

where $r_i(t)$ is the position of the ith particle at time t and the sum is over all such particles of interest. A more manageable field can be formed through coarse-graining. Willard and Chandler [213] used a convolution with the normalized Gaussian functions

$$\phi(r;\zeta) = (2\pi\zeta^2)^{-d/2} exp(-r^2/2\zeta^2) \tag{3.15}$$

where ζ is the coarse-graining length, and d stands for dimensionality. Hence, applied to $\rho(r,t)$ we have the coarse grained density field

$$\overline{\rho}(r,t) = \sum_i \phi(|r - r_i(t)|;\zeta) \tag{3.16}$$

The choice of ζ will depend upon the physical conditions under consideration. Consequently, the instantaneous interface is defined as

$$\overline{\rho}(s,t) = c \tag{3.17}$$

where c is a constant. In other words, the instantaneous interface is defined as points in space where the coarse-grained density has the value c. For molecular configurations, equation 3.17 can be solved quickly through interpolation on a spatial grid. Figure 3.22 [213] illustrates what is found for a slab of liquid water at conditions of water-vapour coexistence. Half of the bulk density $c = 0.016$ Å^{-3} was used as constant and $\zeta = 2.4$ Å for the coarse graining. As described in [213] the time evolution of the liquid-vapour instantaneous interface can then be used for estimation of the surface tension. For that, the power spectrum of the instantaneous interface must be obtained. Hence, the Fourier transform $\tilde{h}(k)$ is obtained from the instantaneous interface configuration $h(x,y)$ according to [213]. The power spectrum of the simulated instantaneous interface is in good agreement with the capillary-wave theory prediction ($\langle|\tilde{h}(k)|^2\rangle \approx 1/\beta\gamma k^2$) for wave-vectors smaller than $\approx 2\pi/9$Å. For larger wave-vectors, the power spectrum is sensitive to molecular detail, hence to the coarse graining factor ζ. Fitting the data of the SPC/E simulation yields $\gamma = 62.0 \pm 0.5$ mJ/m^2 which is in reasonable agreement with the experimental value of 72 mJ/m^2 and simulated values 63.6 ± 1.5 mJ/m^2 [215].

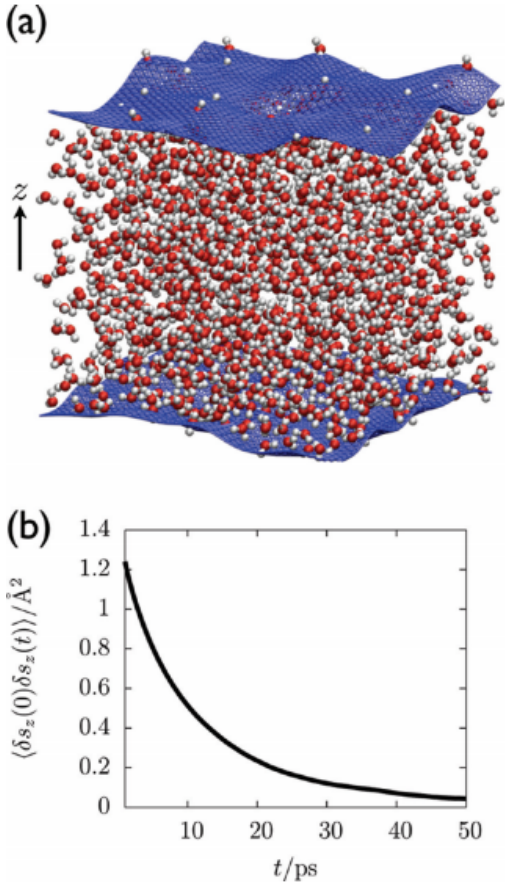


Figure 3.22 – Snapshot of a slab of liquid water with the instantaneous interface s rendered as a blue mesh on the upper and lower phase boundary. The slab is periodically replicated in the horizontal directions. (b) The time correlation function governing the spatial fluctuations in the intrinsic interface s. Here, angle brackets represent an equilibrium average and $\delta s_z(t) = (s(t) \cdot \hat{z} - \langle s \rangle \cdot \hat{z}$, where $s(t)$ is the position of the interface at time t and $\hat{z}$ is the unit vector in the z direction (as is indicated in panel a).

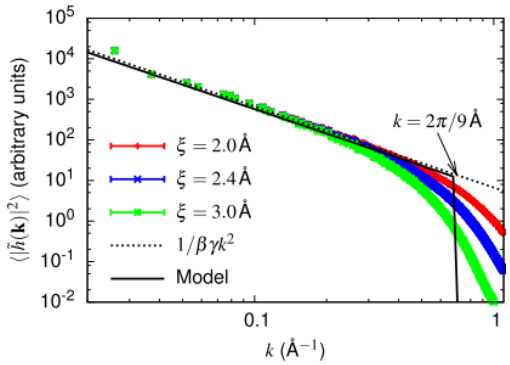
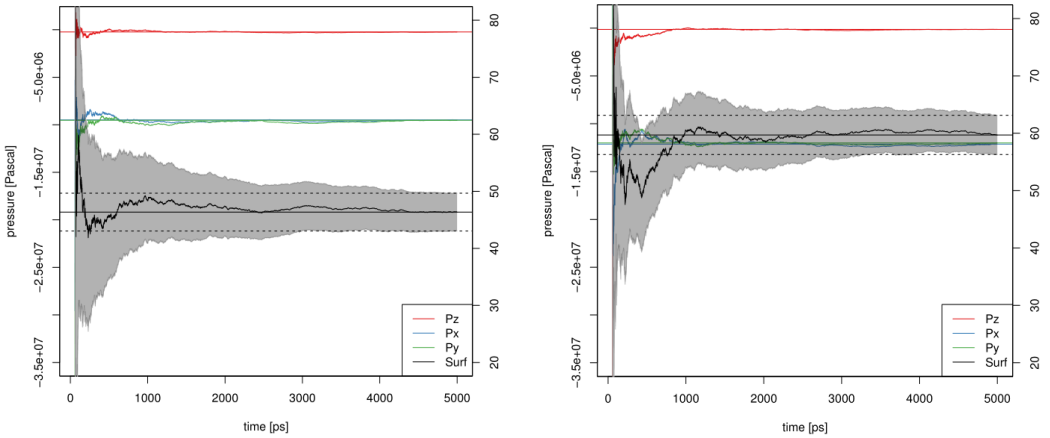


Figure 3.23 – Power spectrum of the interface with different coarse graining ζ compared to the capillary-wave model.

3.8.4.3 Results from PM6 and DFTB simulations



(a) TIP3P simulation

(b) SPC/E simulation

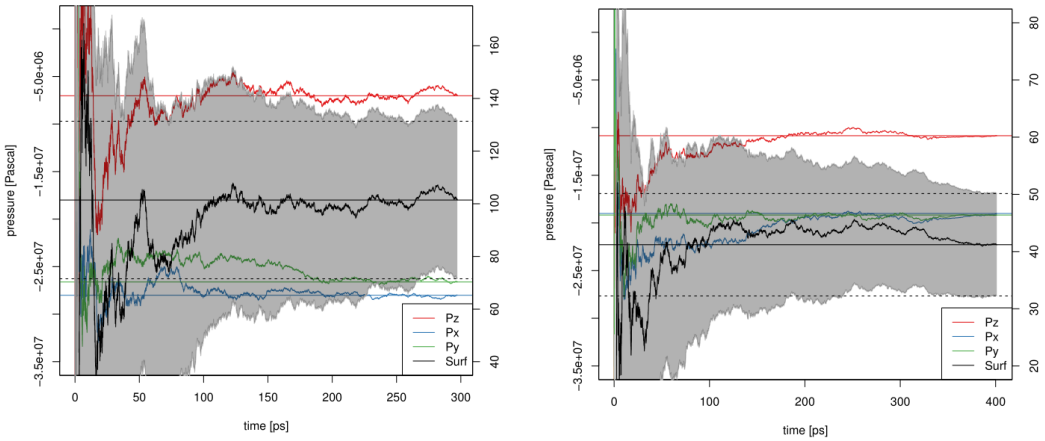
Figure 3.24 – Accumulated average values of the normal P_z and tangential (P_x and P_y) components of the pressure tensor as obtained from direct simulation of the vapour-liquid interface for the TIP3P and SPC/E model of water at a temperature of T = 300 K. The accumulated average value of the surface tension as obtained from the pressure tensor with Equation 3.12 is also presented with corresponding standard deviation.

A first investigation of the classical force field approaches was carried out. Hence in Figure 3.24a and 3.24b the normal (red) and tangential components (blue and green) of the pressure tensor is plotted over time for the slab simulation at T = 300 K for the TIP3P forcefield. The accumulated running averages are shown in the same colors respectively and both later pressure components P_x and P_y converge as expected. The calculated surface tension from Equation 3.13 is plotted as running average (black solid line) with according standard deviation (gray area) and final error estimation (black dotted line). Vega et al [215] stated that a total simulation time of 1.5-2 ns are needed to provide accurate values and Chen et al [216] even used longer trajectories (5 ns) for their analysis. However, the calculated standard deviations of 2 mN/m [216] are in good agreement with our obtained standard deviations of 2.32 mN/m. As from the plot in Figure 3.24a can be interfered, even though it takes long simulation times to get better convergence on the surface tension, to obtain values in the boundaries of the error it appears that shorter trajectories are already sufficient. Therefore, starting from an equilibrated slab it can be assumed that $>$ 1 ns already provide values within the final standard deviation. The same analysis was carried out for the PM6 and DFTB3 simulations and are presented in Figure 3.25a and 3.25b. The final value of the surface tension for the DFTB3 seems comparable to the value obtained from the classical approach with TIP3P. Both, force field and DFTB3 methods underestimate the experimental reference value of 71.97mN/m. The results from the PM6 simulations indicate an overestimation of the surface tension although it is difficult to conclude since the standard deviation is still high. The resulting Table 3.14 summarizes the finding of the classical force-field approaches (TIP3P and

SPC/E) and the semi-empirical methods (PM6 and DFTB3).

Method	TIP3P	SPC/E	PM6	DFTB3
Surface tension γ [mN/m]	46.35	59.56	101.4	41.2
standard deviation σ [mN/m]	2.32	3.55	30.5	8.9

Table 3.14 – Surface tension from mechanical route



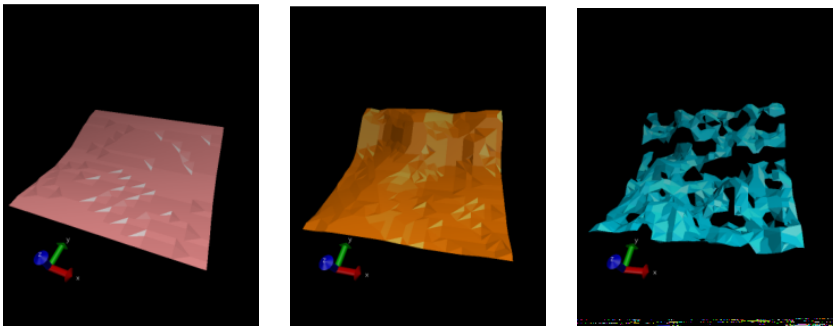
(a) Average values of the Pressure components for the PM6$_{optim}$ simulations.

(b) Average values of the Pressure components for the DFTB3$_{optim}$ simulations.

Figure 3.25 – The normal P_z and tangential (P_x and P_y) components of the pressure tensor shown with according surface tension and standard deviation as obtained from Equation 3.12.

The calculation of the surface tension from fluctuations of the instantaneous interface were conducted with a coarse graining of $\zeta = 1.2$. An analysis as shown in Table 3.15 of different ζ's and a visual inspection of the single water molecule resulted in the decision to use the value of 1.2 for ζ.

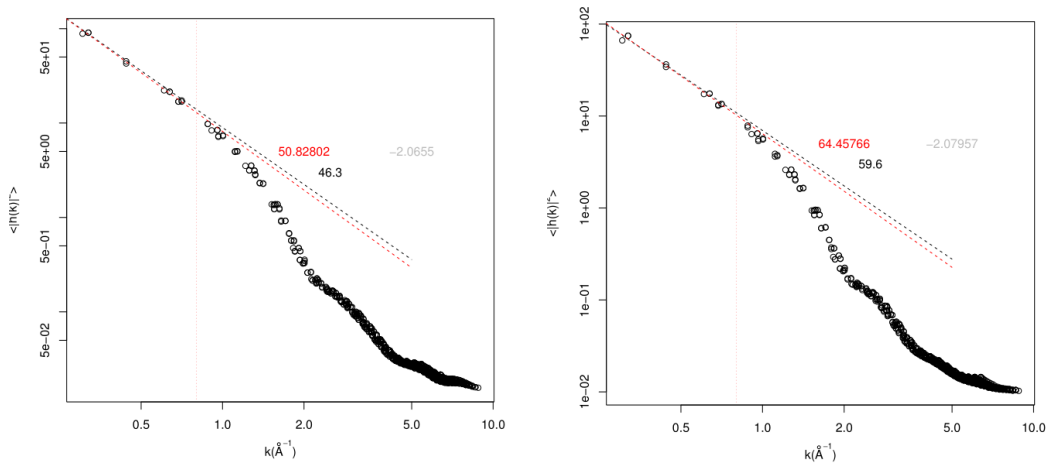
The power spectrum was fitted according to the relation $< |\tilde{h}(k)|^2 > \; 1/\beta\gamma k^2$ for wave-vectors smaller than 0.8 [Å^{-1}]. Where $\beta = 1/(k_b \cdot T)$ and γ is the resulting surface tension. The results for the TIP3P liquid-vapour interface simulation are shown in Figure 3.26a. Where the black dotted line represents the surface tension obtained from the mechanical route (46.3 mN/m) and the red dotted line is the linear regression of the power spectrum resulting in a value of 50.83 mN/m. The conducted analysis and the resulting value are in reasonable agreement with the mechanical route and hence also in the standard deviation for the surface tension as evaluated earlier. An additional affirmation of a good fit is the value for the slope of the curve which turns out to be -2.066 and hence sufficiently close to the expected value of -2.0 of the underlying theory. The regression



$\zeta = 6$	$\zeta = 3$	$\zeta = 1$	$\zeta = 0.5$

Table 3.15 – Different coarse graining ζ

was carried out up to a wave-vector length of k = 0.8 and it is crucial to mention that the fit is very sensitive to that value. In principle only the linear regime of the plot should be fitted and therefore an intended smaller value is used.



(a) Power spectrum of the TIP3P force-field. (b) Power spectrum of the SPC/E force-field.

Figure 3.26 – Power spectrum of the TIP3P interface with a coarse graining of $\zeta = 1.2$ and different wave-vectors k. The black dotted line is the reference value 46.3 mN/m obtained from the mechanical route and red the fitted regression line and the resulting surface tension 50.83 mN/m. In light gray the number from the linear regression is plotted which corresponds quite well to the theoretical value of -2 and hence the resulting value of the surface tension can be assumed to be reasonable.

Method	TIP3P	SPC/E	DFTB3
Surface tension γ [mN/m]	50.82	64.46	27.6

Table 3.16 – Surface tension from instantaneous fluctuations

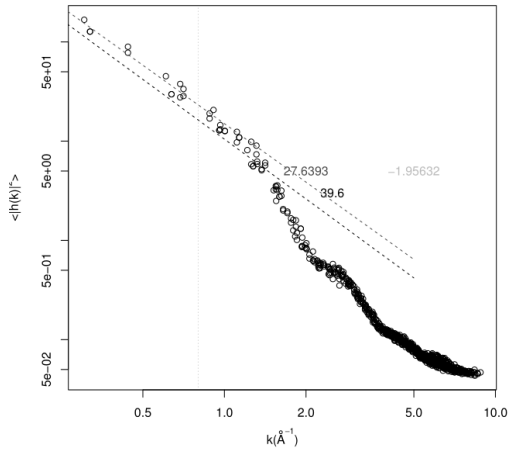


Figure 3.27 – Surface tension from fluctuation for the γ-corrected DFTB3 method

3.8.5 Diffusion

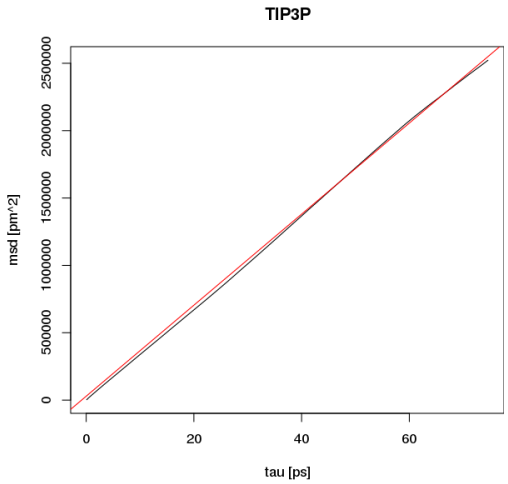
Diffusion is defined by Collins Essential English Dictionary [217] as the random thermal motion of atoms and molecules in gases, liquids, and some solids. The self-diffusion coefficient is the average speed that a specific water molecule holds, in order to diffuse in liquid water. The term 'self' is used to distinguish from 'bulk' diffusion and relates to the progress of distinguishable particles. D_s is measured by incoherent quasi-elastic neutron scattering.
In this study, the self-diffusion coefficient of water was calculated using the Einstein equation [218].

$$D_s = lim_{t\rightarrow \mathrm{inf}} \left[\frac{\langle |r(t) - r(0)|^2 \rangle}{6t} \right] \tag{3.18}$$

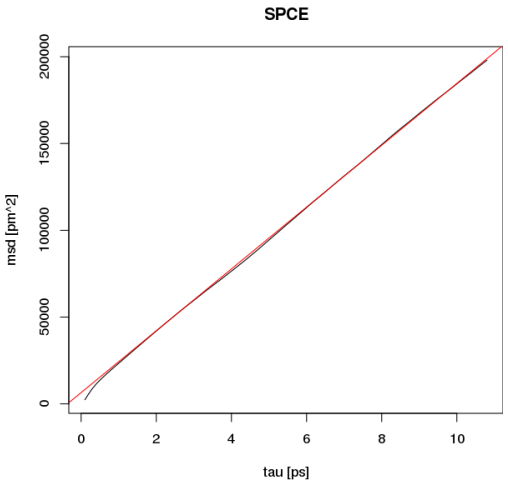
where $r(t)$ is the position vector of the center of mass at time t, and the brackets $\langle \cdots \rangle$ denote an average over both time origins and individual water molecules. The mean square displacement (MSD) refers to the numerator and the experimental self-diffusion value for water, as determined by [219] is $2.3 \pm 15\% \cdot 10^{-9} m^2 s^{-1}$. D was calculated by fitting a straight line to the MSD at intermediate times to avoid contamination from the ballistic regime at smaller times and the lower statistics region at the end of the trajectory.

3.8.5.1 Results from PM6 and DFTB simulations

Resulting MSD plots are shown in Figure 3.28a to 3.30b. In Table 3.17, all self-diffusion coefficients obtained from the NVE simulations are summarized. As expected for the classical force-field approaches, the TIP3P water model overestimates the self-diffusion as well as the PM6$_{optim}$ method. However, the PM6$_{optim}$ parameter set is still in better agreement than the classical calculations. SPC/E is known for performing very

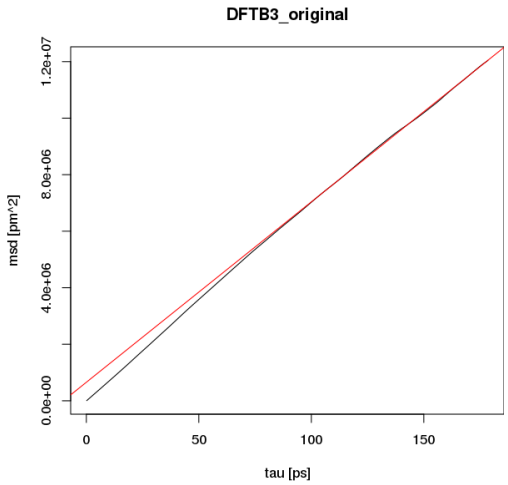


(a) Mean square displacement as a function of time for the TIP3P force field

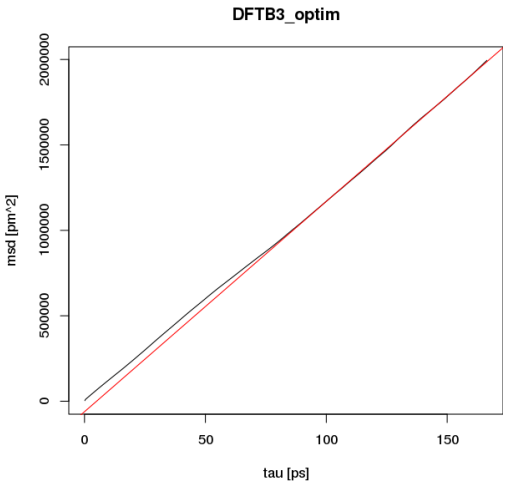


(b) Mean square displacement as a function of time for the SPC/E force field

Figure 3.28 – Mean square displacement and corresponding fitting for the determination of the self-diffusion coefficient for the classical approaches.

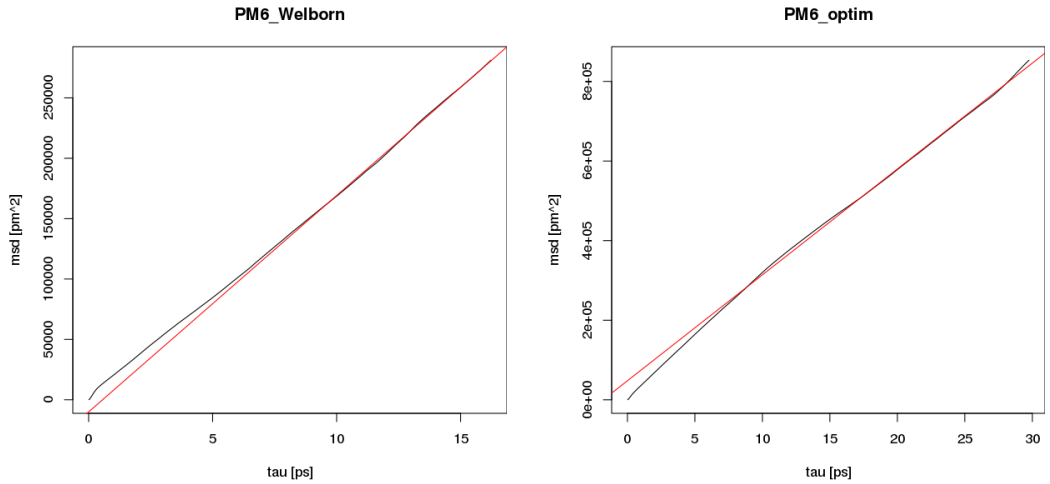


(a) Mean square displacement as a function of time for the original DFTB3 method



(b) Mean square displacement as a function of time for the γ corrected DFTB3 method

Figure 3.29 – Mean square displacement and corresponding fitting for the determination of the self-diffusion coefficient for the both DFTB3 methods.



(a) Mean square displacement as a function of time for the PM6 method with the parameters from Welborn et al [199]

(b) Mean square displacement as a function of time for the PM6 method with optimized parameters

Figure 3.30 – Mean square displacement and corresponding fitting for the determination of the self-diffusion coefficient for both PM6 methods.

well for the diffusion coefficient and the value is indeed very close to the experimental measurement. Furthermore, the parameter set from Welborn et al [199] for the PM6 method demonstrates very good agreement to experiment. The γ-corrected DFTB3 exhibits major improvements over the standard DFTB3 results and reports values also in good accordance with the experimental value.

	TIP3P	SPC/E	DFTB3_original	DFTB3_final	PM6_Welborn	PM6_optim	Experiment
D [10^-9 m^2/s]	5.6319	2.9690	10.6392	2.0513	2.9867	4.4335	2.2260

Table 3.17 – Different self-diffusion values from the NVE simulations of bulk water

4 Summary and Outlook

4.1 Summary

The main purpose of this work was to establish a more systematic approach to parameter optimization in computational chemistry where the subjective weighting of objectives is not part of the optimization process itself but the decision making and selection of the final solution is carried out a posteriori. This separation introduces more transparency and might help to further unleash the potential of all the computational chemistry methods from ab-initio to semi-empirical and force field approaches.
The first part of the thesis was dedicated to elaborating the differences of single-objective optimization and multi-objective optimization. Thus, the theoretical concepts of parameter optimization are reviewed and it is shown that the multi-objective optimization can not be seen as simple extension of the single-objective case. It is included that the usual scalarization techniques to linearly combine different objectives into a single-objective problem by weighting renders the approach unfeasible because some functions can not be added up properly with constant weights. Additionally, not all solutions of the Pareto-front can be recovered with the preference-based approaches especially when the Pareto-set is non-convex. The weighting in the classical algorithms is specially delicate since a priori knowledge is required and it is not clear how the normalization (to render the objectives dimensionless) and the according weighting should be carried out. The multi-objective approach presented in this work demonstrate a systematic way to conduct parameter optimization where more than one objective is involved. This procedure does not introduce any ambiguity or bias and therefore could improve many aspects of parameter optimization in computational chemistry.
Further, a short introduction to Evolution Strategy (ES) algorithms is given. The basic concept is a repeated process of stochastic variations followed by selection, inspired by the principles of biological evolution. This concept is applied to parameter optimization and the Covariance Matrix Adaptation (CMA) algorithm is introduced. The CMA process operates with a covariance where new trial solutions are generated and by accepting the offspring (new solutions) the covariance and mean are adapted. This mechanism

should drive the entire process towards a minimum solution. In principle, the number of offspring can be chosen but allowing only one offspring per parent results in a (1+1) ES selection scheme. The elitist (1+1)-CMA-ES is then presented and outlined as basis for the resulting multi-objective optimization algorithm. For the multi-objective optimization case, a population of λ-(1+1)-CMA-ES instances are maintained. From the population offspring a selection is carried out where solutions are compared over their level of non-dominance and crowding-distance (a measure to achieve a good spread of the Pareto front). Again, the selection process should drive the optimization run towards the Pareto front and recover all feasible trade-off solutions.
The second part of the thesis reviews the common methods in semi-empirical quantum chemistry. The aim is to give a general introduction on semi-empirical methods and the essentials of modern semi-empirical molecular orbital theory. The grounds of semi-empirical molecular orbital theory is based on two approximations: self-consistent field (SCF) and linear combination of atomic orbitals (LCAO). The former is also known as mean-field approximation where the one-electron Hamiltonians are used to solve for the wave-function of the system. The latter is a proposal by Huckel that the molecular orbitals can be represented as a linear combination of atomic orbitals. These simplifications are then used to solve the Schrödinger equation. Finally two models, the Neglect of Diatomic Differential Overlap (NDDO) and Density Functional Tight Binding (DFTB) are outlined and all the necessary parameters explained. There are many methods based on the NDDO model and they differ manly in the core-core repulsion terms. One of the most prominent (PM6) is then used for later re-parametrization.
In the last part of this work we apply multi-objective parameter optimization to the semi-empirical quantum chemistry methods and intended to answer several questions. First, if it is possible to reproduce bulk properties from cluster reference data. Hence, a review of available water cluster databases is carried out and with help of quantum cluster equilibration (QCE) a final selection obtained. Indeed, with the small selection of 8 water clusters the PM6 method could be re-parametrized to give good bulk properties. This finding is relevant not only because it demonstrates the possibility that bulk behaviour can be recovered from cluster but also the results might contribute to the understanding and structure of liquid water. Second, we wanted to show that the underlying model of NDDO is capable of capturing important properties such as hydrogen bonds and van der Waals forces. Again, with the successful re-parametrization of PM6 we could show that this method provides enough flexibility to describe such phenomena even though we allowed the individual parameters only to change ±10% from the original values. We could show that there is no need for additional dispersion or hydrogen bonding correction as was proposed in the past. Therefore, we advocate a rigorous re-parametrization before enhancing a method and thus explore the boundaries of the underlying model first.

4.2 Outlook

The work presented in this thesis provides the ground work for a number of new areas of research. An immediate area of interest is the selection of the reference set in more detail. There is consensus one aspect of water's molecular structure sets it apart from most other liquids is the hydrogen bonds. However, based on its ability to form up to 4 hydrogen bonds it is clear that the water dimer itself would not be representative for water. But that is where the consensus ends. The standard picture of liquid water [220] posits that each molecule of H_2O is, on average, bonded to four others in a tetrahedral motif. This would result in a repeated, constantly reorganizing three-dimensional network extending through the liquid. Hence, a reference set of tetrahedral coordinated water would represent best that picture. In 2004 an different view on water emerged [220] advocating that molecules in liquid water bind on average to just two others, forming chains and rings. From their point of view, a reference set with mainly 2-fold coordinated water molecule would make more sense. In this work, we relied on the selection that was carried out with the quantum cluster equilibration model and adds up to a view that suggest a muddle of two different structures. Employing a bulk reference set for multi-objective optimization with objectives such as error in forces and energies or dipole moments and comparing the results (the final Pareto-front) with the one obtained from smaller water cluster sets could help to improve our understanding of liquid water. In other words, simply comparing optimization runs from reference bulk sets (from ab-initio calculations) with cluster structures could shed some light on which structures are well described and might help unveil irrelevant clusters and therefore add some insight to the discussion on coordination of liquid water. Thus, the multi-objective approach could contribute towards an understanding and answering the important question of what clusters can tell us about the bulk.

Concerning the parameter optimization in the specific case of PM6 one important input to the optimization process was the boundaries that were set only ±10 % of the original values. One should investigate the effect of larger bounds. Giving the optimization routine more flexibility should further improve the description of water at ambient condition. However, caution is advised when setting the boundaries since setting the limits too wide would break the transferability and invalid some properties that were accounted for in the past. Therefore, a closer look at parameter boundaries and other properties such as ionization potential should be carried out.

Of course other objectives for the optimization process could be imposed besides the ones employed in the present work. Hence not only the effects of the parameter bounds but also the consequences of different objectives should be investigated. One could include the ionization potential directly into the optimization process or include some sort of electro-static potential fitting where highly accurate ab-initio reference sets would be needed. Additionally, non-physical aspects such as SCF cycles or run times could be used to optimize parameter to faster molecular dynamics simulations. As we have seen, for the DFTB3 γ-function optimization, the charge and water monomer geometry have

not changed substantially and hence an additional run without those objective might improve results. In general, it would be very interesting to apply the multi-objective framework to re-parametrize the entire DFTB3 repulsive terms. Hence, this would be more concordant with our approach of global optimization and that corrections (such as the γ-function correction) should only be applied after a rigorous re-parametrization of the underlying method.

A designated area that would benefit from the multi-objective approach is the classical force field development. Accurate force fields are needed to obtain meaningful results from molecular dynamics simulations and recent developments are trying to incorporate quantum mechanics explicitly to construct the inter-atomic potentials. Usually, either experimental data such as enthalpy of vaporization, enthalpy of sublimation, dipole moments and various spectroscopic parameters or ab-initio calculations are used as reference data. Multi-objective optimization with highly accurate ab-initio reference data could be used to improve the classical force field approach. Even short molecular dynamics calculations could be used in the optimization process itself to determine properties such as self-diffusion or radial distribution functions which in turn can be used as independent objectives. Needless to say that this is only possible because molecular dynamics with force fields are very fast and therefore simply obtained. Hence, because multi-objective optimization provides a great systematic framework with a wide range of flexibility it would be a great application to improve the classical force fields and push the accuracy limits even further.

A significant share of molecular calculations in quantum chemistry is based on density functional theory (DFT). The success of DFT can be attributed to maintaining a good accuracy while the computational cost is low. One of the major technical aspects of DFT calculations is the basis set used to solve the Kohn-Sham equations. There exists a wide range of used functional forms, including Gaussian functions, Slater functions, plane waves, wavelets, numerical basis function, and many more. For each of these functional forms, there are schemes to increase the size, and hence the accuracy of the basis. Often, there is no perfect recipe to do so, and different schemes will be used to obtain, for example, good total energies, geometries, interaction energies, or special electronic properties. Therefore, multi-objective optimization could also prove useful in the field of basis set optimization where simultaneously several objectives are optimized. The gain from optimal small basis sets is enormous since at fixed system size, the density matrix update procedure scale at least quadratically or even cubically with number of basis functions. Another advantage of the multi-objective approach for basis set optimization is that additional non-physical objectives can be added. Such objectives might be the condition number of the overlap matrix since this quantity governs computational cost of diagonalization or total number of SCF cycles. Hence performance based descriptors can be added directly in the parameter optimization process. Even more, the Pareto set for a given reference data and basis set size could be stored in a database. These tables then can be accessed by the end-user and by providing the desired objectives and requested accuracy, a resulting basis set is returned with the required demands.

As a final remark, it can be noted that, due to the general aspects of the multi-objective optimization algorithm, the possible applications are virtually infinite. Although, in many cases there exist more specialized and problem specific algorithms that also include gradient information or other available constraints or resources for the single-objective case. However, once one deals with more than one objective it is crucial to be aware of the limitation of the scalarization approach and with the present thesis we expect to pave the way for a more systematic take on parameter optimization in computational chemistry.

Bibliography

[1] K. Ramachandran, G. Deepa, and K. Namboori. "Computational chemistry and molecular modeling: principles and applications." Springer Science & Business Media, **2008**.

[2] E. Drexler. "Engines of Creation. The Coming Era of Nanotechnology." New York: Anchor Books, **1986**.

[3] E. Schrödinger. "Quantisierung als Eigenwertproblem." Annalen der Physik, 385(13):437–490, **1926**.

[4] D. R. Hartree. "The wave mechanics of an atom with a non-Coulomb central field. Part I. Theory and methods." In "Mathematical Proceedings of the Cambridge Philosophical Society," volume 24, pages 89–110. Cambridge University Press, **1928**.

[5] J. C. Slater. "A simplification of the Hartree-Fock method." Physical Review, 81(3):385, **1951**.

[6] G. Segal. "Semiempirical Methods of Electronic Structure Calculation: Part B: Applications.", volume 8. Springer Science & Business Media, **2012**.

[7] F. Jensen. "Introduction to Computational Chemistry." John Wiley & Sons, **2013**.

[8] C. Dykstra, G. Frenking, K. Kim, and G. Scuseria. "Theory and Applications of Computational Chemistry: the first forty years." Elsevier, **2005**.

[9] C. J. Cramer. "Essentials of Computational Chemistry: theories and models." John Wiley & Sons, **2013**.

[10] U. Burkert, N. L. Allinger, et al. "Molecular mechanics.", volume 177. American Chemical Society Washington, DC, **1982**.

[11] V. Gold, K. Loening, A. McNaught, and P. Shemi. "IUPAC Compendium of Chemical Terminology." Blackwell Science, Oxford, **1997**.

[12] F. Weinhold. "Quantum cluster equilibrium theory of liquids: General theory and computer implementation." The Journal of Chemical Physics, 109(2):367–372, **1998**.

[13] S. Lehmann, C. Spickermann, and B. Kirchner. "Quantum Cluster Equilibrium Theory Applied in Hydrogen Bond Number Studies of Water. 1. Assessment of the Quantum Cluster Equilibrium Model for Liquid Water." Journal of Chemical Theory and Computation, 5(6):1640–1649, **2009**.

[14] N. Hansen and A. Ostermeier. "Adapting arbitrary normal mutation distributions in evolution strategies: The covariance matrix adaptation." In "Evolutionary Computation, 1996., Proceedings of IEEE International Conference on," pages 312–317. IEEE, **1996**.

[15] T. Weise. "Global optimization algorithms-theory and application." Citeseer, **2009**.

[16] K. Deb and D. Kalyanmoy. "Multi-Objective Optimization Using Evolutionary Algorithms." John Wiley & Sons, **2001**.

[17] L. Pauling. "The Nature of the Chemical Bond.", volume 3. Cornell University Press, **1960**.

[18] W. Michiels, E. Aarts, and J. Korst. "Theoretical aspects of local search." Springer Science & Business Media, **2007**.

[19] S. S. Skiena. "The algorithm design manual: Text.", volume 1. Springer Science & Business Media, **1998**.

[20] Kendall Atkinson. "An Introduction to Numerical Analysis, 2nd Edition." John Wiley & Sons, **1989**.

[21] W. K. Hastings. "Monte Carlo sampling methods using Markov chains and their applications." Biometrika, 57(1):97–109, **1970**.

[22] K. E. Batcher. "Design of a massively parallel processor." Computers, IEEE Transactions on, 100(9):836–840, **1980**.

[23] Swiss National Supercomputing Centre. Website: `http://www.cscs.ch` [2015, October 6] , **2015**.

[24] T. Bäck and H.-P. Schwefel. "An overview of evolutionary algorithms for parameter optimization." Evolutionary computation, 1(1):1–23, **1993**.

[25] L. A. Zadeh. "What is soft computing?" Soft computing, 1(1):1–1, **1997**.

[26] R. A. Brooks. "Intelligence without representation." Artificial intelligence, 47(1):139–159, **1991**.

[27] F. Glover. "Tabu search-part 1." ORSA Journal on computing, 1(3):190–206, **1989**.

[28] W. J. Baumol and P. Hall. "Economic theory and operations analysis." Prentice-Hall Englewood Cliffs, NJ, **1977**.

[29] M. Broadie. "Computing efficient frontiers using estimated parameters." Annals of Operations Research, 45(1):21–58, **1993**.

[30] A. E. Bryson. "Applied optimal control: optimization, estimation and control." CRC Press, **1975**.

[31] P. Y. Papalambros and D. J. Wilde. "Principles of optimal design: modeling and computation." Cambridge University Press, **2000**.

[32] S. Das and B. K. Panigrahi. "Multi-Objective Evolutionary Algorithms." Encyclopedia of artificial intelligence, 3:1145–1151, **2009**.

[33] J. J. Stewart. "Optimization of parameters for semiempirical methods V: modification of NDDO approximations and application to 70 elements." Journal of Molecular modeling, 13(12):1173–1213, **2007**.

[34] Y. G. Evtushenko and M. A. Posypkin. "Method of non-uniform coverages to solve the multicriteria optimization problems with guaranteed accuracy." Automation and Remote Control, 75(6):1025–1040, **2014**.

[35] V. Podinovskii and V. Nogin. "Pareto-optimal solutions of multicriteria problems." Moscow: Sci, **1982**.

[36] L. Amoroso. "Vilfredo Pareto." Econometrica: Journal of the Econometric Society, pages 1–21, **1938**.

[37] H. Fang, Q. Wang, Y.-C. Tu, and M. F. Horstemeyer. "An efficient non-dominated sorting method for evolutionary algorithms." Evolutionary computation, 16(3):355–384, **2008**.

[38] L. Ding, S. Zeng, and L. Kang. "A fast algorithm on finding the non-dominated set in multi-objective optimization." In "Evolutionary Computation, 2003. CEC'03. The 2003 Congress on," volume 4, pages 2565–2571. IEEE, **2003**.

[39] J. Du and Z. Cai. "A Sorting Based Algorithm for Finding a Non-dominated Set in Multi-objective Optimization." In "Natural Computation, 2007. ICNC 2007. Third International Conference on," volume 4, pages 436–440. IEEE, **2007**.

[40] K. Mishra and S. Harit. "A fast algorithm for finding the non dominated set in multi-objective optimization." International Journal of Computer Applications, 1(25):3539, **2010**.

[41] J. L. Cohon. "Multicriteria programming: Brief review and application." Design optimization, pages 163–191, **1985**.

[42] C.-L. Hwang, S. R. Paidy, K. Yoon, and A. S. M. Masud. "Mathematical programming with multiple objectives: A tutorial." Computers & Operations Research, 7(1):5–31, **1980**.

[43] K. Miettinen. "Nonlinear multiobjective optimization.", volume 12. Springer Science & Business Media, **1998**.

[44] Y. Y. Haimes, L. S. Lasdon, and D. A. Wismer. "On a bicriterion formulation of the problems of integrated system identification and system optimization." IEEE Transactions on Systems Man and Cybernetics, SMC-1(1):296–297, **1971**.

[45] R. Enkhbat, J. Guddat, and A. Chinchuluun. "Parametric Multiobjective Optimization." In "Pareto Optimality, Game Theory And Equilibria," pages 529–538. Springer, **2008**.

[46] K. Q. Zhao and X. M. Yang. "E-Benson proper efficiency in vector optimization." Optimization, 64(4):739–752, **2015**.

[47] W. Zhang and T. Gao. "A min–max method with adaptive weightings for uniformly spaced Pareto optimum points." Computers & structures, 84(28):1760–1769, **2006**.

[48] I. Kaliszewski. "A modified weighted Tchebycheff metric for multiple objective programming." Computers & operations research, 14(4):315–323, **1987**.

[49] J. Jahn. "Scalarization in multi objective optimization." Springer, **1985**.

[50] M. Dorigo, M. Birattari, C. Blum, M. Clerc, T. Stützle, and A. Winfield. "Ant Colony Optimization and Swarm Intelligence: 6th International Conference, ANTS 2008, Brussels, Belgium, September 22-24, 2008, Proceedings.", volume 5217. Springer, **2008**.

[51] D. J. Cavicchio. "Adaptive search using simulated evolution." Technical report, National Institutes of Health, **1970**.

[52] X. Jin and R. G. Reynolds. "Using knowledge-based evolutionary computation to solve nonlinear constraint optimization problems: a cultural algorithm approach." In "Evolutionary Computation, 1999. CEC 99. Proceedings of the 1999 Congress on," volume 3. IEEE, **1999**.

[53] X.-S. Yang. "Nature-inspired metaheuristic algorithms." Luniver press, **2010**.

[54] T. Bäck. "Evolutionary algorithms in theory and practice: evolution strategies, evolutionary programming, genetic algorithms." Oxford University Press, **1996**.

[55] E. Zitzler and L. Thiele. "Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach." Evolutionary Computation, IEEE transactions on, 3(4):257–271, **1999**.

[56] C. A. C. Coello, D. A. Van Veldhuizen, and G. B. Lamont. "Evolutionary algorithms for solving multi-objective problems.", volume 242. Springer, **2002**.

[57] C. M. Fonseca and P. J. Fleming. "An overview of evolutionary algorithms in multiobjective optimization." Evolutionary computation, 3(1):1–16, **1995**.

[58] Z. Michalewicz and M. Schoenauer. "Evolutionary algorithms for constrained parameter optimization problems." Evolutionary computation, 4(1):1–32, **1996**.

[59] T. Back, D. B. Fogel, and Z. Michalewicz. "Handbook of evolutionary computation." IOP Publishing Ltd., **1997**.

[60] T. Bäck, D. B. Fogel, and Z. Michalewicz. "Evolutionary computation 1: Basic algorithms and operators.", volume 1. CRC Press, **2000**.

[61] N. Hansen and A. Ostermeier. "Completely derandomized self-adaptation in evolution strategies." Evolutionary computation, 9(2):159–195, **2001**.

[62] C. L. Müller and I. F. Sbalzarini. "Gaussian Adaptation revisited–an entropic view on covariance matrix adaptation." In "Applications of Evolutionary Computation," pages 432–441. Springer, **2010**.

[63] N. Hansen. "Invariance, self-adaptation and correlated mutations in evolution strategies." In "Parallel Problem Solving from Nature PPSN VI," pages 355–364. Springer, **2000**.

[64] N. Hansen, S. D. Müller, and P. Koumoutsakos. "Reducing the time complexity of the derandomized evolution strategy with covariance matrix adaptation (CMA-ES)." Evolutionary Computation, 11(1):1–18, **2003**.

[65] G. Jastrebski, D. V. Arnold, et al. "Improving evolution strategies through active covariance matrix adaptation." In "Evolutionary Computation, 2006. CEC 2006. IEEE Congress on," pages 2814–2821. IEEE, **2006**.

[66] H.-G. Beyer and B. Sendhoff. "Covariance matrix adaptation revisited - the CMSA evolution strategy -." In "Parallel Problem Solving from Nature–PPSN X," pages 123–132. Springer, **2008**.

[67] N. Hansen. "The CMA evolution strategy: a comparing review." In "Towards a new evolutionary computation," pages 75–102. Springer, **2006**.

[68] N. Hansen and S. Kern. "Evaluating the CMA evolution strategy on multimodal test functions." In "Parallel problem solving from nature-PPSN VIII," pages 282–291. Springer, **2004**.

[69] CMA-ES. "CMA-ES — Wikipedia, The Free Encyclopedia.", **2015**. [2015, September 29].

[70] C. Igel, N. Hansen, and S. Roth. "Covariance matrix adaptation for multi-objective optimization." Evolutionary computation, 15(1):1–28, **2007**.

[71] A. Ostermeier, A. Gawelczyk, and N. Hansen. "A derandomized approach to self-adaptation of evolution strategies." Evolutionary Computation, 2(4):369–380, **1994**.

[72] K. Deb and K. Miettinen. "Multiobjective optimization: Interactive and evolutionary approaches.", volume 5252. Springer Science & Business Media, **2008**.

[73] C. L. Müller. "Black-box landscapes: characterization, optimization, sampling, and application to geometric configuration problems." na, **2010**.

[74] A. Konak, D. W. Coit, and A. E. Smith. "Multi-objective optimization using genetic algorithms: A tutorial." Reliability Engineering & System Safety, 91(9):992–1007, **2006**.

[75] G. Rudolph. "Convergence properties of evolutionary algorithms." Kovac, **1997**.

[76] A. Zhou, B.-Y. Qu, H. Li, S.-Z. Zhao, P. N. Suganthan, and Q. Zhang. "Multiobjective evolutionary algorithms: A survey of the state of the art." Swarm and Evolutionary Computation, 1(1):32–49, **2011**.

[77] E. Zitzler, K. Deb, and L. Thiele. "Comparison of multiobjective evolutionary algorithms: Empirical results." Evolutionary computation, 8(2):173–195, **2000**.

[78] L. While, P. Hingston, L. Barone, and S. Huband. "A faster algorithm for calculating hypervolume." Evolutionary Computation, IEEE Transactions on, 10(1):29–38, **2006**.

[79] L. While, L. Bradstreet, L. Barone, and P. Hingston. "Heuristics for optimizing the calculation of hypervolume for multi-objective optimization problems." In "Evolutionary Computation, 2005. The 2005 IEEE Congress on," volume 3, pages 2225–2232. IEEE, **2005**.

[80] O. Kramer. "A review of constraint-handling techniques for evolution strategies." Applied Computational Intelligence and Soft Computing, 2010, **2010**.

[81] C. Delerue and M. Lannoo. "Nanostructures: theory and modelling." Springer Science & Business Media, **2004**.

[82] E. G. Lewars. "Computational Chemistry: introduction to the theory and applications of molecular and quantum mechanics." Springer Science & Business Media, **2010**.

[83] W. Thiel. "Semiempirical methods: current status and perspectives." Tetrahedron, 44(24):7393–7408, **1988**.

[84] N. Otte, M. Scholten, and W. Thiel. "Looking at self-consistent-charge density functional tight binding from a semiempirical perspective." The Journal of Physical Chemistry A, 111(26):5751–5755, **2007**.

[85] M. Elstner. "The SCC-DFTB method and its application to biological systems." Theoretical Chemistry Accounts, 116(1-3):316–325, **2006**.

[86] M. Gaus, Q. Cui, and M. Elstner. "DFTB3: extension of the self-consistent-charge density-functional tight-binding method (SCC-DFTB)." J. Chem. Theory Comput, 7(4):931–948, **2011**.

[87] T. Clark and J. J. Stewart. "MNDO-Like Semiempirical Molecular Orbital Theory and Its Application to Large Systems." Computational methods for large systems: Electronic structure approaches for biotechnology and nanotechnology, pages 259–286, **2011**.

[88] J. A. McCammon and S. C. Harvey. "Dynamics of proteins and nucleic acids." Cambridge University Press, **1988**.

[89] W. L. Jorgensen and J. Tirado-Rives. "Potential energy functions for atomic-level simulations of water and organic and biomolecular systems." Proceedings of the National Academy of Sciences of the United States of America, 102(19):6665–6670, **2005**.

[90] J. N. Israelachvili. "Intermolecular and surface forces: revised third edition." Academic press, **2011**.

[91] P. Koehl, M. Levitt, et al. "A brighter future for protein structure prediction." Nature structural biology, 6:108–111, **1999**.

[92] W. Thiel. "Semiempirical quantum–chemical methods." Wiley Interdisciplinary Reviews: Computational Molecular Science, 4(2):145–157, **2014**.

[93] R. G. Parr. "Density functional theory." Annual Review of Physical Chemistry, 34(1):631–656, **1983**.

[94] E. K. Gross and R. M. Dreizler. "Density functional theory.", volume 337. Springer Science & Business Media, **2013**.

[95] F. P. Boer, M. D. Newton, and W. N. Lipscomb. "Extended HÜckel theory and molecular hartree-fock SCF theory." Proceedings of the National Academy of Sciences of the United States of America, 52(4):890, **1964**.

[96] R. Hoffmann. "An extended Hückel theory. I. hydrocarbons." The Journal of Chemical Physics, 39(6):1397–1412, **1963**.

[97] K. Ohno. "Some remarks on the Pariser-Parr-Pople method." Theoretica chimica acta, 2(3):219–227, **1964**.

[98] I. Fischer-Hjalmars. "Zero Differential Overlap in pi Electron Theories." Advances in quantum chemistry, 2:25, **1965**.

[99] J. A. Pople, D. P. Santry, and G. A. Segal. "Approximate Self-Consistent Molecular Orbital Theory. I. Invariant Procedures." The Journal of Chemical Physics, 43(10):S129–S135, **1965**.

[100] M. Gordon and J. Pople. "Approximate Self-Consistent Molecular-Orbital Theory. VI. INDO Calculated Equilibrium Geometries." The Journal of Chemical Physics, 49(10):4643–4650, **1968**.

[101] H. Hase and A. Schweig. "CNDO/2 (complete neglect of differential overlap)-method for third-row molecules." Theoretica chimica acta, 31(3):215–220, **1973**.

[102] J. Pople, D. Beveridge, and P. Dobosh. "Approximate Self-Consistent Molecular-Orbital Theory. V. Intermediate Neglect of Differential Overlap." The Journal of Chemical Physics, 47(6):2026–2033, **1967**.

[103] R. C. Bingham, M. J. Dewar, and D. H. Lo. "Ground states of molecules. XXV. MINDO/3. Improved version of the MINDO semiempirical SCF-MO method." Journal of the American Chemical Society, 97(6):1285–1293, **1975**.

[104] M. J. Dewar and W. Thiel. "Ground states of molecules. 38. The MNDO method. Approximations and parameters." Journal of the American Chemical Society, 99(15):4899–4907, **1977**.

[105] M. J. Dewar and W. Thiel. "Ground states of molecules. 39. MNDO results for molecules containing hydrogen, carbon, nitrogen, and oxygen." Journal of the American Chemical Society, 99(15):4907–4917, **1977**.

[106] M. J. Dewar, E. G. Zoebisch, E. F. Healy, and J. J. Stewart. "Development and use of quantum mechanical molecular models. 76. AM1: a new general purpose quantum mechanical molecular model." Journal of the American Chemical Society, 107(13):3902–3909, **1985**.

[107] J. J. Stewart. "Optimization of parameters for semiempirical methods I. Method." Journal of Computational Chemistry, 10(2):209–220, **1989**.

[108] D. Nanda and K. Jug. "SINDO1. A semiempirical SCF MO method for molecular binding energy and geometry I. Approximations and parametrization." Theoretica chimica acta, 57(2):95–106, **1980**.

[109] K. Jug, R. Iffert, and J. Schulz. "Development and parametrization of sindo1 for second-row elements." International journal of quantum chemistry, 32(2):265–277, **1987**.

[110] B. Ahlswede and K. Jug. "Consistent modifications of SINDO1: I. approximations and parameters." Journal of Computational Chemistry, 20(6):563–571, **1999**.

[111] J. Ridley and M. Zerner. "An intermediate neglect of differential overlap technique for spectroscopy: pyrrole and the azines." Theoretica chimica acta, 32(2):111–134, **1973**.

[112] A. D. Bacon and M. C. Zerner. "An intermediate neglect of differential overlap theory for transition metal complexes: Fe, Co and Cu chlorides." Theoretica chimica acta, 53(1):21–54, **1979**.

[113] M. J. Dewar and W. Thiel. "A semiempirical model for the two-center repulsion integrals in the NDDO approximation." Theoretica chimica acta, 46(2):89–104, **1977**.

[114] W. Thiel and A. A. Voityuk. "Extension of the MNDO formalism tod orbitals: Integral approximations and preliminary numerical results." Theoretica chimica acta, 81(6):391–404, **1992**.

[115] J. J. Stewart. "Optimization of parameters for semiempirical methods VI: more modifications to the NDDO approximations and re-optimization of parameters." Journal of molecular modeling, 19(1):1–32, **2013**.

[116] M. P. Repasky, J. Chandrasekhar, and W. L. Jorgensen. "PDDG/PM3 and PDDG/MNDO: improved semiempirical methods." Journal of Computational Chemistry, 23(16):1601–1622, **2002**.

[117] M. Kolb and W. Thiel. "Beyond the MNDO model: Methodical considerations and numerical results." Journal of Computational Chemistry, 14(7):775–789, **1993**.

[118] W. Weber and W. Thiel. "Orthogonalization corrections for semiempirical methods." Theoretical Chemistry Accounts, 103(6):495–506, **2000**.

[119] A. F. Oliveira, G. Seifert, T. Heine, and H. A. Duarte. "Density-functional based tight-binding: an approximate DFT method." Journal of the Brazilian Chemical Society, 20(7):1193–1205, **2009**.

[120] M. Elstner, D. Porezag, G. Jungnickel, J. Elsner, M. Haugk, T. Frauenheim, S. Suhai, and G. Seifert. "Self-consistent-charge density-functional tight-binding method for simulations of complex materials properties." Physical Review B, 58(11):7260, **1998**.

[121] H. Ehrenreich and M. H. Cohen. "Self-consistent field approach to the many-electron problem." Physical Review, 115(4):786, **1959**.

[122] T. Clark and R. Koch. "Linear Combination of Atomic Orbitals." In "The Chemist's Electronic Book of Orbitals," pages 5–22. Springer, **1999**.

[123] M. Born and R. Oppenheimer. "Zur Quantentheorie der Molekeln." Annalen der Physik, 389(20):457–484, **1927**.

[124] V. Fock. "Bemerkung zur Quantelung des harmonischen Oszillators im Magnetfeld." Zeitschrift für Physik, 47(5-6):446–448, **1928**.

[125] W. Pauli. "Über den Zusammenhang des Abschlusses der Elektronengruppen im Atom mit der Komplexstruktur der Spektren." Zeitschrift für Physik A Hadrons and Nuclei, 31(1):765–783, **1925**.

[126] J. C. Slater. "The theory of complex spectra." Physical Review, 34(10):1293, **1929**.

[127] E. Hückel. "Quantum-theoretical contributions to the benzene problem. I. The electron configuration of benzene and related compounds." Z. physik, 70(3-4):204–286, **1931**.

[128] O. Sinanoğlu and D. F.-t. Tuan. "Many-Electron Theory of Atoms and Molecules. III. Effect of Correlation on Orbitals." The Journal of Chemical Physics, 38(7):1740–1748, **1963**.

[129] T. Clark and R. Koch. "The Chemist's Electronic Book of Orbitals." Journal fur praktische Chemie, 341:701–701, **1999**.

[130] A. Kramida, Yu. Ralchenko, J. Reader, and and NIST ASD Team. NIST Atomic Spectra Database (ver. 5.2), [Online]. Available: `http://physics.nist.gov/asd` [2015, October 6]. National Institute of Standards and Technology, Gaithersburg, MD., **2014**.

[131] P. J. Linstrom and W. G. Mallard, editors. "NIST Chemistry WebBook, NIST Standard Reference Database Number 69." National Institute of Standards and Technology, Gaithersburg MD, 20899, **June 2005**.

[132] W. C. Martin and A. Musgrove. "Ground levels and ionization energies for the neutral atoms." NIST Physics Laboratory, **1998**.

[133] F. H. Allen. "The Cambridge Structural Database: a quarter of a million crystal structures and rising." Acta Crystallographica Section B: Structural Science, 58(3):380–388, **2002**.

[134] D. F. Shanno. "On broyden-fletcher-goldfarb-shanno method." Journal of Optimization Theory and Applications, 46(1):87–94, **1985**.

[135] K. Y. Burstein and A. N. Isaev. "MNDO calculations on hydrogen bonds. Modified function for core-core repulsion." Theoretica chimica acta, 64(5):397–401, **1984**.

[136] G. I. Csonka and J. G. Ángyán. "The origin of the problems with the PM3 core repulsion function." Journal of Molecular Structure: THEOCHEM, 393(1):31–38, **1997**.

[137] A. A. Voityuk and N. Rösch. "AM1/d parameters for molybdenum." The Journal of Physical Chemistry A, 104(17):4089–4094, **2000**.

[138] J. J. Stewart. "Optimization of parameters for semiempirical methods 2. Applications." Technical report, DTIC Document, **1989**.

[139] J. J. Stewart. "Optimization of parameters for semiempirical methods. III Extension of PM3 to Be, Mg, Zn, Ga, Ge, As, Se, Cd, In, Sn, Sb, Te, Hg, Tl, Pb, and Bi." Journal of Computational Chemistry, 12(3):320–341, **1991**.

[140] J. J. Stewart. "Optimization of parameters for semiempirical methods IV: extension of MNDO, AM1, and PM3 to more main group elements." Journal of Molecular Modeling, 10(2):155–164, **2004**.

[141] W. Kohn and L. J. Sham. "Self-consistent equations including exchange and correlation effects." Physical Review, 140(4A):A1133, **1965**.

[142] G. Seifert, D. Porezag, and T. Frauenheim. "Calculations of molecules, clusters, and solids with a simplified LCAO-DFT-LDA scheme." International journal of quantum chemistry, 58(2):185–192, **1996**.

[143] D. Porezag, T. Frauenheim, T. Köhler, G. Seifert, and R. Kaschner. "Construction of tight-binding-like potentials on the basis of density-functional theory: Application to carbon." Physical Review B, 51(19):12947, **1995**.

[144] M. Elstner. "SCC-DFTB: What is the proper degree of self-consistency?" The Journal of Physical Chemistry A, 111(26):5614–5621, **2007**.

[145] M. Gaus, A. Goez, and M. Elstner. "Parametrization and benchmark of DFTB3 for organic molecules." Journal of Chemical Theory and Computation, 9(1):338–354, **2012**.

[146] Y. Yang, H. Yu, D. York, Q. Cui, and M. Elstner. "Extension of the self-consistent-charge density-functional tight-binding method: third-order expansion of the density functional theory total energy and introduction of a modified effective coulomb interaction." The Journal of Physical Chemistry A, 111(42):10861–10873, **2007**.

[147] H. Eschrig. "Optimized LCAO method and the electronic structure of extended systems." Ser.

[148] M. Gaus, C.-P. Chou, H. Witek, and M. Elstner. "Automatized Parametrization of SCC-DFTB Repulsive Potentials: Application to Hydrocarbons." The Journal of Physical Chemistry A, 113(43):11866–11881, **2009**.

[149] CP2K developer group. Available: `http://www.cp2k.org` [2015, October 6]. , **2015**.

[150] J. Rezác, J. Fanfrlik, D. Salahub, and P. Hobza. "Semiempirical quantum chemical PM6 method augmented by dispersion and H-bonding correction terms reliably describes various types of noncovalent complexes." Journal of Chemical Theory and Computation, 5(7):1749–1760, **2009**.

[151] CP2K developer group and Y. Misteli. Available: `https://github.com/misteliy/cp2k-mo-cma` [2015, October 6]. , **2015**.

[152] A. K. Jain and R. C. Dubes. "Algorithms for clustering data." Prentice-Hall, **1988**.

[153] J. J. Stewart. "Application of the PM6 method to modeling proteins." Journal of molecular modeling, 15(7):765–805, **2009**.

[154] J. J. Stewart. "Application of the PM6 method to modeling the solid state." Journal of molecular modeling, 14(6):499–535, **2008**.

[155] Z. Marković, D. Milenković, J. Jorović, J. M. D. Marković, V. Stepanić, B. Lučić, and D. Amić. "PM6 and DFT study of free radical scavenging activity of morin." Food chemistry, 134(4):1754–1760, **2012**.

[156] A. Alparone, V. Librando, and Z. Minniti. "Validation of semiempirical PM6 method for the prediction of molecular properties of polycyclic aromatic hydrocarbons and fullerenes." Chemical Physics Letters, 460(1):151–154, **2008**.

[157] A. d. S. Gonçalves, T. C. França, J. D. Figueroa-Villar, and P. G. Pascutti. "Conformational analysis of toxogonine, TMB-4 and HI-6 using PM6 and RM1 methods." Journal of the Brazilian Chemical Society, 21(1):179–184, **2010**.

[158] G. Murdachaew, C. J. Mundy, G. K. Schenter, T. Laino, and J. Hutter. "Semiempirical Self-Consistent Polarization Description of Bulk Water, the Liquid- Vapor Interface, and Cubic Ice." The Journal of Physical Chemistry A, 115(23):6046–6053, **2011**.

[159] E. A. Coutsias, C. Seok, and K. A. Dill. "Using quaternions to calculate RMSD." Journal of Computational Chemistry, 25(15):1849–1857, **2004**.

[160] J. Cioslowski. "Encyclopedia of Computational Chemistry." Wiley, New York, **1998**.

[161] F. Martin and H. Zipse. "Charge distribution in the water molecule—a comparison of methods." Journal of Computational Chemistry, 26(1):97–105, **2005**.

[162] W. Benedict, N. Gailar, and E. K. Plyler. "Rotation-Vibration Spectra of Deuterated Water Vapor." The Journal of Chemical Physics, 24(6):1139–1165, **1956**.

[163] B. T. Darling and D. M. Dennison. "The water vapor molecule." Physical Review, 57(2):128, **1940**.

[164] S. A. Clough, Y. Beers, G. P. Klein, and L. S. Rothman. "Dipole moment of water from Stark measurements of H2O, HDO, and D2O." The Journal of Chemical Physics, 59(5):2254–2259, **1973**.

[165] V. S. Bryantsev, M. S. Diallo, A. C. van Duin, and W. A. Goddard III. "Evaluation of B3LYP, X3LYP, and M06-class density functionals for predicting the binding energies of neutral, protonated, and deprotonated water clusters." Journal of Chemical Theory and Computation, 5(4):1016–1026, **2009**.

[166] R. M. Shields, B. Temelso, K. A. Archer, T. E. Morrell, and G. C. Shields. "Accurate Predictions of Water Cluster Formation,(H2O) n= 2-10." The Journal of Physical Chemistry A, 114(43):11725–11737, **2010**.

[167] F. Weinhold. "Nature of H-bonding in clusters, liquids, and enzymes: an ab initio, natural bond orbital perspective." Journal of Molecular Structure: THEOCHEM, 398:181–197, **1997**.

[168] R. Ludwig. "Water: from clusters to the bulk." Angewandte Chemie, International Edition, 40(10):1808–1827, **2001**.

[169] R. Ahlrichs, M. Bär, M. Häser, H. Horn, and C. Kölmel. "Electronic structure calculations on workstation computers: The program system turbomole." Chemical Physics Letters, 162(3):165–169, **1989**.

[170] M. D. Ben. "Efficient Non-Local Dynamical Electron Correlation for Condensed Matter Simulations." Ph.D. thesis, University of Zurich, **2014**.

[171] P. Politzer, J. S. Murray, and P. Lane. "Electrostatic potentials and covalent radii." Journal of Computational Chemistry, 24(4):505–511, **2003**.

[172] W. Klopper, J. Van Duijneveldt-Van De Rijdt, and F. Van Duijneveldt. "Computational determination of equilibrium geometry and dissociation energy of the water dimer." Physical Chemistry Chemical Physics, 2(10):2227–2234, **2000**.

[173] P. Goyal, M. Elstner, and Q. Cui. "Application of the SCC-DFTB method to neutral and protonated water clusters and bulk water." The Journal of Physical Chemistry B, 115(20):6790–6805, **2011**.

[174] P. Goyal, H.-J. Qian, S. Irle, X. Lu, D. Roston, T. Mori, M. Elstner, and Q. Cui. "Molecular simulation of water and hydration effects in different environments: Challenges and developments for DFTB based models." The Journal of Physical Chemistry B, 118(38):11007–11027, **2014**.

[175] T. H. Choi, R. Liang, C. M. Maupin, and G. A. Voth. "Application of the SCC-DFTB method to hydroxide water clusters and aqueous hydroxide solutions." The Journal of Physical Chemistry B, 117(17):5165–5179, **2013**.

[176] B. Guillot. "A reappraisal of what we have learnt during three decades of computer simulations on water." Journal of Molecular Liquids, 101(1):219–260, **2002**.

[177] J. R. Errington and P. G. Debenedetti. "Relationship between structural order and the anomalies of liquid water." Nature, 409(6818):318–321, **2001**.

[178] H. E. Stanley, S. V. Buldyrev, and N. Giovambattista. "Static heterogeneities in liquid water." Physica A: Statistical Mechanics and its Applications, 342(1):40–47, **2004**.

[179] A. Nilsson and L. G. Pettersson. "Perspective on the structure of liquid water." Chemical Physics, 389(1):1–34, **2011**.

[180] I.-F. W. Kuo, C. J. Mundy, M. J. McGrath, J. I. Siepmann, J. VandeVondele, M. Sprik, J. Hutter, B. Chen, M. L. Klein, F. Mohamed, et al. "Liquid water from first principles: Investigation of different sampling approaches." The Journal of Physical Chemistry B, 108(34):12990–12998, **2004**.

[181] P. Wernet, D. Nordlund, U. Bergmann, M. Cavalleri, M. Odelius, H. Ogasawara, L. Näslund, T. Hirsch, L. Ojamäe, P. Glatzel, et al. "The structure of the first coordination shell in liquid water." Science, 304(5673):995–999, **2004**.

[182] K. Laasonen, M. Sprik, M. Parrinello, and R. Car. "Ab-initio liquid water." The Journal of Chemical Physics, 99(11):9080–9089, **1993**.

[183] S. Izvekov and G. A. Voth. "Car–Parrinello molecular dynamics simulation of liquid water: New results." The Journal of Chemical Physics, 116(23):10372–10376, **2002**.

[184] J. C. Grossman, E. Schwegler, E. W. Draeger, F. Gygi, and G. Galli. "Towards an assessment of the accuracy of density functional theory for first principles simulations of water." The Journal of Chemical Physics, 120(1):300–311, **2004**.

[185] I.-C. Lin, A. P. Seitsonen, I. Tavernelli, and U. Rothlisberger. "Structure and Dynamics of Liquid Water from ab Initio Molecular Dynamics Comparison of BLYP, PBE, and revPBE Density Functionals with and without van der Waals Corrections." Journal of Chemical Theory and Computation, 8(10):3902–3910, **2012**.

[186] J. Schmidt, J. VandeVondele, I.-F. W. Kuo, D. Sebastiani, J. I. Siepmann, J. Hutter, and C. J. Mundy. "Isobaric- isothermal molecular dynamics simulations utilizing density functional theory: An assessment of the structure and density of water at near-ambient conditions." The Journal of Physical Chemistry B, 113(35):11959–11964, **2009**.

[187] J. VandeVondele, F. Mohamed, M. Krack, J. Hutter, M. Sprik, and M. Parrinello. "The influence of temperature and density functional models in ab initio molecular dynamics simulation of liquid water." The Journal of Chemical Physics, 122(1):014515, **2005**.

[188] M. J. McGrath, J. I. Siepmann, I.-F. W. Kuo, C. J. Mundy, J. VandeVondele, J. Hutter, F. Mohamed, and M. Krack. "Isobaric–Isothermal Monte Carlo Simulations from First Principles: Application to Liquid Water at Ambient Conditions." ChemPhysChem, 6(9):1894–1901, **2005**.

[189] M. C. Zerner. "Semiempirical molecular orbital methods." Reviews in Computational Chemistry, 2:313–365, **1991**.

[190] H. Liu, M. Elstner, E. Kaxiras, T. Frauenheim, J. Hermans, and W. Yang. "Quantum mechanics simulation of protein dynamics on long timescale." Proteins: Structure, Function, and Bioinformatics, 44(4):484–489, **2001**.

[191] M. Elstner, Q. Cui, P. Munih, E. Kaxiras, T. Frauenheim, and M. Karplus. "Modeling zinc in biomolecules with the self consistent charge-density functional tight binding (SCC-DFTB) method: Applications to structural and energetic analysis." Journal of Computational Chemistry, 24(5):565–581, **2003**.

[192] D. J. Price and C. L. Brooks III. "A modified TIP3P water potential for simulation with Ewald summation." The Journal of Chemical Physics, 121(20):10096–10103, **2004**.

[193] H. Berendsen, J. Grigera, and T. Straatsma. "The missing term in effective pair potentials." Journal of Physical Chemistry, 91(24):6269–6271, **1987**.

[194] M. Del Ben, M. Schonherr, J. Hutter, and J. VandeVondele. "Bulk liquid water at ambient temperature and pressure from MP2 theory." The Journal of Physical Chemistry Letters, 4(21):3753–3759, **2013**.

[195] G. A. Jeffrey and G. A. Jeffrey. "An introduction to hydrogen bonding.", volume 12. Oxford University Press New York, **1997**.

[196] A. Soper and M. Phillips. "A new determination of the structure of water at 25 C." Chemical Physics, 107(1):47–60, **1986**.

[197] A. Soper, F. Bruni, and M. Ricci. "Site–site pair correlation functions of water from 25 to 400 C: Revised analysis of new and old diffraction data." The Journal of Chemical Physics, 106(1):247–254, **1997**.

[198] M. Schoenherr. "High Performance Tree Monte Carlo applied to solid and liquid water." Ph.D. thesis, University of Zurich, **2013**.

[199] M. Welborn, J. Chen, L.-P. Wang, and T. Van Voorhis. "Why many semiempirical molecular orbital theories fail for liquid water and how to fix them." Journal of Computational Chemistry, 36(12):934–939, **2015**.

[200] M. Brehm and B. Kirchner. "TRAVIS-a free analyzer and visualizer for Monte Carlo and molecular dynamics trajectories." Journal of Chemical Information and Modeling, 51(8):2007–2023, **2011**.

[201] R. S. Taylor, L. X. Dang, and B. C. Garrett. "Molecular dynamics simulations of the liquid/vapor interface of SPC/E water." The Journal of Physical Chemistry, 100(28):11720–11725, **1996**.

[202] S. E. Feller, R. W. Pastor, A. Rojnuckarin, S. Bogusz, and B. R. Brooks. "Effect of electrostatic force truncation on interfacial and transport properties of water." The Journal of Physical Chemistry, 100(42):17011–17020, **1996**.

[203] D. M. Huang, P. L. Geissler, and D. Chandler. "Scaling of hydrophobic solvation free energies." The Journal of Physical Chemistry B, 105(28):6704–6709, **2001**.

[204] J. Rivera, M. Predota, A. Chialvo, and P. Cummings. "Vapor–liquid equilibrium simulations of the SCPDP model of water." Chemical Physics Letters, 357(3):189–194, **2002**.

[205] A. Wynveen and F. Bresme. "Interactions of polarizable media in water: A molecular dynamics approach." The Journal of Chemical Physics, 124(10):104502, **2006**.

[206] A. E. Ismail, G. S. Grest, and M. J. Stevens. "Capillary waves at the liquid-vapor interface and the surface tension of water." The Journal of Chemical Physics, 125(1):014702, **2006**.

[207] B. Shi, S. Sinha, and V. K. Dhir. "Molecular dynamics simulation of the density and surface tension of water by particle-particle particle-mesh method." The Journal of Chemical Physics, 124(20):204715, **2006**.

[208] V. V. Zakharov, E. N. Brodskaya, and A. Laaksonen. "Surface tension of water droplets: A molecular dynamics study of model and size dependencies." The Journal of Chemical Physics, 107(24):10675–10683, **1997**.

[209] B. C. Garrett, G. K. Schenter, and A. Morita. "Molecular simulations of the transport of molecules across the liquid/vapor interface of water." Chemical Reviews, 106(4):1355–1374, **2006**.

[210] I. W. Kuo, C. J. Mundy, B. L. Eggimann, M. J. McGrath, J. I. Siepmann, B. Chen, J. Vieceli, and D. J. Tobias. "Structure and Dynamics of the Aqueous Liquid-Vapor Interface: A Comprehensive Particle-Based Simulation Study." The Journal of Physical Chemistry B, 110(8):3738–3746, **2006**.

[211] J. Rowlinson and B. Widom. "Molecular theory of capillarity. The international series of monographs on chemistry." Clarendon Press, Oxford, **1982**.

[212] J. Vrabec, G. K. Kedia, G. Fuchs, and H. Hasse. "Comprehensive study of the vapour–liquid coexistence of the truncated and shifted Lennard–Jones fluid including planar and spherical interface properties." Molecular physics, 104(09):1509–1527, **2006**.

[213] A. P. Willard and D. Chandler. "Instantaneous liquid interfaces." The Journal of Physical Chemistry B, 114(5):1954–1958, **2010**.

[214] G. J. Gloor, G. Jackson, F. J. Blas, and E. de Miguel. "Test-area simulation method for the direct determination of the interfacial tension of systems with continuous or discontinuous potentials." The Journal of Chemical Physics, 123(13):134703, **2005**.

[215] C. Vega and E. De Miguel. "Surface tension of the most popular models of water by using the test-area simulation method." The Journal of Chemical Physics, 126(15):154707, **2007**.

[216] F. Chen and P. E. Smith. "Simulated surface tensions of common water models." The Journal of Chemical Physics, 126(22):221101, **2007**.

[217] E. Collins. "Essential English Dictionary.", **1989**.

[218] D. Tildesley and M. Allen. "Computer simulation of liquids." Clarendon, Oxford, **1987**.

[219] K. Krynicki, C. D. Green, and D. W. Sawyer. "Pressure and temperature dependence of self-diffusion in water." Faraday Discussions of the Chemical Society, 66:199–208, **1978**.

[220] P. Ball. "Water: water—an enduring mystery." Nature, 452(7185):291–292, **2008**.